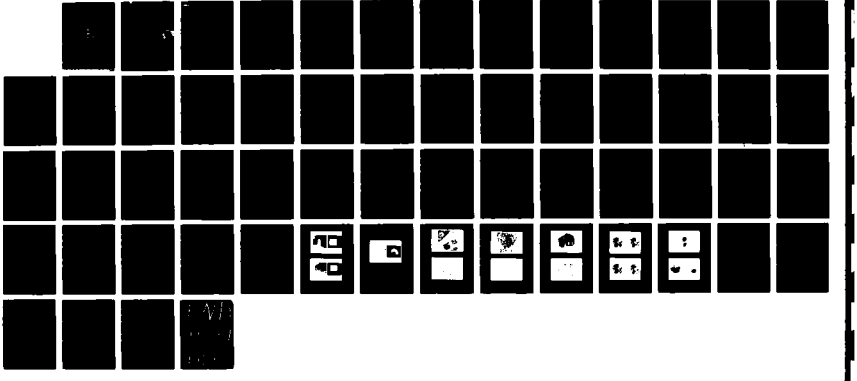
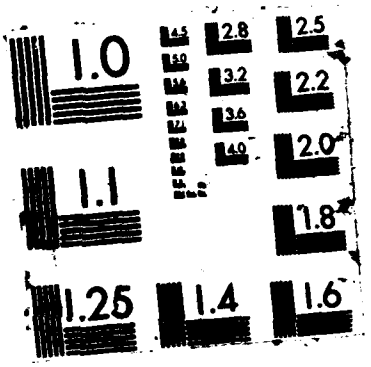
AD-A185 719 A UNIFIED MATHEMATICAL APPROACH TO IMAGE ANALYSIS(U) 1/1
BROWN UNIV PROVIDENCE RI DIV OF APPLIED MATHEMATICS
S GEMAN ET AL. 31 AUG 87 ARO-20551.4-MA
UNCLASSIFIED DAAG29-83-K-0116 F/G 17/11 NL




1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8

2

AD-A185 719

REPORT DOCUMENTATION PAGE

1b. RESTRICTIVE MARKINGS

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution unlimited.

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
ARO 20551.4-MA

6a. NAME OF PERFORMING ORGANIZATION
Brown University

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
U. S. Army Research Office

6c. ADDRESS (City, State, and ZIP Code)
Division of Applied Mathematics
Box F
Providence, RI 02912

7b. ADDRESS (City, State, and ZIP Code)
P. O. Box 12211
Research Triangle Park, NC 27709-2211

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
U. S. Army Research Office

8b. OFFICE SYMBOL (If applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DAAG29-83-K-0116

8c. ADDRESS (City, State, and ZIP Code)
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
A Unified Mathematical Approach to Image Analysis

12. PERSONAL AUTHOR(S)
Geman, Stuart; Geman, Donald (U.Mass); Grenander, Ulf; McClure, Donald E.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Technical	FROM 1Sep83 TO 30Apr87	1987 August 31	55

16. SUPPLEMENTARY NOTATION
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Image processing, Bayesian models, Algorithms, Segmentation, Boundary detection; tomography; Global image analysis.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes a unified approach to the mathematical modelling and statistical analysis of pictures. The results reported here were derived from the research project "Nonparametric Estimation by the Model of Sieves" supported by the U.S. Army Research Office, Contract DAAG29-83-K-0116 to Brown University. Early in the project, the investigations focused on the estimation of (properties of) digitized images; this focus was a natural extension of (i) our earlier work on the method of sieves of regularizing statistical estimates of infinite dimensional objects and (ii) the groundwork that had been laid for General Pattern Theory. The present report presents the Bayesian paradigm for mathematical modeling and analysis of digital images and then describes four instances of the paradigm in detail. Directions for ongoing and future research are also indicated.

Keywords: Image processing; Algorithms; Segmentation;

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ARO 20561.4-MA

A UNIFIED MATHEMATICAL APPROACH TO IMAGE ANALYSIS

FINAL TECHNICAL REPORT

Stuart Geman
Donald Geman
Ulf Grenander
Donald E. McClure

August 1987

U.S. ARMY RESEARCH OFFICE

CONTRACT NUMBER DAAG29-83-K-0116

Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

A UNIFIED MATHEMATICAL APPROACH TO IMAGE ANALYSIS

ABSTRACT. This report describes a unified approach to the mathematical modelling and statistical analysis of pictures. The results reported here were derived from the research project "Nonparametric Estimation by the Method of Sieves" supported by the U.S. Army Research Office, Contract DAAG29-83-K-0116 to Brown University. Early in the project, the investigations focussed on the estimation of (properties of) digitized images; this focus was a natural extension of (i) our earlier work on the method of sieves for regularizing statistical estimates of infinite dimensional objects and (ii) the groundwork that had been laid for General Pattern Theory. The present report presents the Bayesian paradigm for mathematical modeling and analysis of digital images and then describes four instances of the paradigm in detail. Directions for ongoing and future research are also indicated.

1. INTRODUCTION

Computational image analysis encompasses a variety of applications involving a sensing device, a computer, and software for restoring and possibly interpreting the sensed data. Most commonly, visible light is sensed by a video camera and converted to an array of measured light intensities, each element corresponding to a small patch in the scene (a picture element, or pixel). The image is thereby digitized, and this format is suitable for computer analysis. In some applications, the sensing mechanism responds to other forms of light, such as in infrared imaging where the camera is tuned to the invisible part of the spectrum neighboring the color red. Infrared light is emitted in proportion to temperature, and thus infrared imaging is suitable for detecting and analyzing the temperature profile of a scene. Applications include automated inspection in industrial settings, medical diagnosis, and targeting and tracking of military objects. In single photon emission tomography, as a diagnostic tool, individual photons, emitted from a radiopharmaceutical (isotope combined with a suitable pharmaceutical) are detected. The objective is to reconstruct the distribution of isotope density inside the body from the externally-collected counts. Depending on the pharmaceutical, the isotope density may correspond to local blood flow (perfusion) or local metabolic activity. Other applications of computer vision include satellite imaging for weather and crop yield prediction, radar imaging in military applications, ultrasonic imaging for industrial inspection and a host of medical applications, and there is a growing role for video imaging in robotics.

The variety of applications has yielded an equal variety of algorithms for restoration

and interpretation. Unfortunately, few general principals have emerged and no common foundation has been layed. Algorithms are by and large *ad hoc*; they are typically dedicated to a single application, and often critically tuned to the particulars of the environment (lighting, weather conditions, magnification, and so-on) in which they are implemented. It is likely that a coherent theoretical framework would support more robust and more powerful algorithms. A well-studied candidate is regularization theory (see [45] and [52], and references therein), which has been successfully applied to a variety of vision tasks. We have been exploring a related approach based upon probabilistic image models, well-defined principals of inference, and a Monte Carlo computation theory. Exploiting this framework, we have recently obtained encouraging results in several areas of application, including tomography, texture analysis, and scene segmentation.

In the section that follows (§2) we lay out, briefly, our paradigm in its general formulation. Then, in each of §3, §4, and §5, we discuss an area of application, to texture segmentation and classification, boundary detection, and single photon emission tomography, respectively. In each application section we summarize our previous work and identify directions for further research. In §6 we describe extensions of the methodology to more ambitious, high-level, vision problems. This entails extending our probability image models to accommodate more global, geometric, image attributes.

2. BAYESIAN PARADIGM

In real scenes, neighboring pixels typically have similar intensities; boundaries are usually smooth and often straight; textures, although sometimes random locally, define spatially homogeneous regions; and objects, such as grass, tree trunks, branches and leaves, have preferred relations and orientations. Our approach to picture processing is to articulate such regularities mathematically, and then to exploit them in a statistical framework to make inferences. The regularities are rarely deterministic; instead, they describe correlations and likelihoods. This leads us to the Bayesian formulation, in which prior expectations are formally represented by a probability distribution. Thus we design a distribution (a prior) on relevant scene attributes to capture the tendencies and constraints that characterize the scenes of interest. Picture processing is then guided by this prior distribution, which, if properly conceived, enormously limits the plausible restorations and interpretations.

The approach involves five steps, which we shall briefly review here (see [22] and [33] for more details). This will define the general framework, and then, in the following three sections, we will concentrate on the analysis of texture, boundary detection, and tomography as illustrative applications.

IMAGE MODEL. This is a probability distribution on relevant image attributes. Both for reasons of mathematical and computational convenience, we use *Markov random fields* (MRFs) as prior probability distributions. Let us suppose that we index all of the relevant attributes by the index set S. The set S is application specific. It typically includes indices for each of the pixels (about 512×512 in the usual video digitization) and may have other indices for such attributes as boundary elements, texture labels, object labels and so-on. Associated with each site $s \in S$ is a real-valued random variable X_s, representing the state of the corresponding attribute. Thus X_s may be the measured intensity at pixel s (typically, $X_s \in \{0,...255\}$), or simply 1 or 0 as a boundary element at location s is present or absent.

The kind of knowledge we represent by the prior distribution is usually local, which is to say that we articulate regularities in terms of small local collections of variables. In the end, this leads to a distribution on $X = \{X_s\}_{s \in S}$ with a more or less local neighborhood structure (again, we refer to [22] and [33] for details). Specifically, our priors are Markov random fields: there exists a (symmetric) *neighborhood relation* $G = \{G_s\}_{s \in S}$, wherein $G_s \subseteq S$ is the set of neighbors of s, such that

$$\Pi(X_s = x_s | X_r = x_r, r \in S, r \neq s) = \Pi(X_s = x_s | X_r = x_r, r \in G_s)$$

$\Pi(a|b)$ is conditional probability, and, by convention, $s \notin G_s$. G symmetric means $s \in G_r \Leftrightarrow r \in G_s$. (Here, we assume that the range of the random vector X is discrete; there are obvious modifications for the continuous or mixed case.)

It is well known, and very convenient, that a distribution Π defines a MRF on S with neighborhood relation G if and only if it is Gibbs with respect to the same graph, (S, G). The latter means that Π has the representation

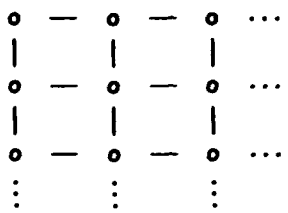
$$\Pi(x) = \frac{1}{z} \exp\{-U(x)\} \tag{2.1}$$

where

$$U(x) = \sum_{c \in C} V_c(x) \tag{2.2}$$

C is the collection of all cliques in (S, G) (a clique is a collection c of sites such that every two sites in c are neighbors), and $V_c(x)$ is a function depending only on $\{x_s\}_{s \in c}$. U is known as the energy, and has the intuitive property that the low energy states are the more likely states under Π. The normalizing constant, z, is known as the partition function. The Gibbs distribution arises in statistical mechanics as the equilibrium distribution of a system with energy function U.

As a simple example (too simple to be of much use for real pictures) suppose the pixel intensities are known, a priori, to be one of two levels, minus one (black) or plus one (white). Let S be the $N \times N$ square lattice, and let G be the neighborhood system that corresponds to nearest horizontal and vertical neighbors:



For picture processing, think of N as typically 512. Suppose that the only relevant regularity is that neighboring pixels tend to have the same intensities. An energy consistent with this regularity is the Ising potential:

$$U(x) = -\beta \sum_{[s,t]} x_s x_t, \qquad \beta > 0$$

where $\sum_{[s,t]}$ means summation over all neighboring pairs $s, t \in S$. The minimum of U is achieved when $x_s = x_t, \quad \forall s, t \in S$. Under (2.1), the likely pictures are therefore the ones that respect our prior expectations; they segment into regions of constant intensities. The larger β, the larger the typical region. Later we will discuss the issue of *estimating* model parameters such as β. (With energy function above, Π in (2.1) is called the Ising model. It models the equilibrium distribution of the spin states of the atoms in a ferromagnet. These spins tend to align themselves with each other and hence the favored configurations contain connected regions of constant spins.)

One very good reason for using MRF priors is their Gibbs representations. Gibbs distributions are characterized by their energy functions, and these are more convenient and intuitive for modelling than working directly with probabilities. See, for example, [20], [22], [24], [33], and [45] for many more examples, and §3, §4, and §5 below for more complex and useful MRF models.

DEGRADATION MODEL. The image model is a distribution $\Pi(\cdot)$ on the vector of image attributes $X = \{X_s\}_{s \in S}$. *By design*, the components of this vector contain all of the relevant information for the image processing task at hand. Hence, the goal is to estimate X. This estimation will be based upon partial or corrupted observations, and based upon the image model, i.e., the prior distribution. In emission tomography, X represents the spatial distribution of isotope in a target region of the body. What is actually observed is

a collection of photon counts whose probability law is Poisson, with a mean function that is an attenuated Radon transform of X. In the texture labelling problem, X is the pixel intensity array combined with a corresponding array of texture labels. Each label gives the texture type of the associated pixel. The observation is only partial: we observe the pixels, which are just the digitized picture, but not the labels. The purpose is then to estimate the labels from the picture.

The observations are related to the image process X by a *degradation model.* This models the relation between X and the *observation process*, say $Y = \{Y_s\}_{s \in T}$. For texture analysis, we will define $X = (X^P, X^L)$, where X^P is the usual grey-level pixel intensity process, and X^L is an associated array of texture labels. The observed picture is just X^P, and hence $Y = X^P$: the degradation is a projection. More typically, the degradation involves a random component, as in the tomography setting where the observations are Poisson variables whose means are related to the image process X. A simpler, and widely studied (if unrealistic), example is additive white noise. Let $X = \{X_s\}_{s \in S}$ be just the basic pixel process. In this case $T = S$, and for each $s \in S$ we observe

$$Y_s = X_s + \eta_s$$

where, for example, $\{\eta_s\}_{s \in S}$ is Gaussian with independent components, having means 0 and variances σ^2, and $\{\eta_s\}$ is independent of the X-process.

Formally, the degradation model is a conditional probability distribution, or density, for Y given X: $\Pi(y|x)$. If the degradation is just additive white noise, as in the above example, then

$$\Pi(y|x) = \left(\frac{1}{2\pi\sigma^2}\right)^{|S|/2} \exp\{-\frac{1}{2\sigma^2}\sum_{s \in S}(y_s - x_s)^2\}$$

For labelling textures, the degradation is deterministic; $\Pi(y|x)$ is concentrated on $y = x^P$, where $x = (x^P, x^L)$ has both pixel and label components.

POSTERIOR DISTRIBUTION. This is the conditional distribution on the image process X given the observation process Y. This posterior or *a posteriori* distribution contains the information relevant to the image restoration or image analysis task. Given an observation $Y = y$, and assuming the image model ($\Pi(x)$) and degradation model ($\Pi(y|x)$), the posterior distribution reveals the likely and unlikely states of the "true" (unobserved) image X. Having constructed X to contain all relevant image attributes, such as locations of boundaries, labels of objects or textures, and so on, the posterior distribution comes to play the fundamental role in our approach to image processing.

The posterior distribution is easily derived from Bayes' rule

$$\Pi(x|y) = \frac{\Pi(y|x)\Pi(x)}{\Pi(y)}$$

The denominator, $\Pi(y)$, is difficult to evaluate. It derives from the prior and degradation models by integration: $\Pi(y) = \int \Pi(y|x)\Pi(dx)$, but the formula is computationally intractable. Happily, our analysis of the posterior distribution will require only *ratios*, not absolute probabilities. Since y is fixed by observation, $1/\Pi(y)$ is a constant that can be ignored (see paragraph below on Computing).

As an example we consider the simple Ising model prior, with observations corrupted by additive white noise. Then

$$\Pi(x) = \frac{1}{z}\exp\{\beta \sum_{[s,t]} x_s x_t\}$$

and

$$\Pi(y|x) = \left(\frac{1}{2\pi\sigma^2}\right)^{|S|/2} \exp\{-\frac{1}{2\sigma^2}\sum_{s\in S}(y_s - x_s)^2\}$$

The posterior distribution is then

$$\Pi(x|y) = \frac{1}{z_p}\exp\{\beta \sum_{[s,t]} x_s x_t - \frac{1}{2\sigma^2}\sum_{s\in S}(y_s - x_s)^2\}$$

We denote by z_p the normalizing constant for the posterior distribution. Of course, it depends on y, but the latter is fixed. Notice that the posterior distribution is again a MRF. In the case of additive white noise, the neighborhood system of the posterior distribution is that of the prior, and hence local.

For a wide class of useful degradation models, including combinations of blur, additive or multiplicative colored noise, and a variety of nonlinear transformations, the posterior distribution is a MRF with a more or less local graph structure. This is convenient for our computational schemes, as we shall see shortly. We should note, however, that exceptions occur. Indeed, nonlocal graph structures that incorporate long-range interactions are useful and are completely consistent with the Bayesian paradigm. In tomography, for example, the posterior distribution is associated with a highly non-local graph. This particular situation incurs a high computational cost (see [24] and §5 below for more details) as a consequence of each site in the graph having a high degree.

ESTIMATING THE IMAGE. In our framework, image processing amounts to choosing a particular image x, given an observation $Y = y$. One choice is the maximum a posteriori, or MAP estimate:

$$\text{choose } x \text{ to maximize } \Pi(x|y)$$

The MAP estimate chooses the most likely x, given the observation. In many applications, our goal is to identify the MAP estimate, or a suitable approximation. Often, though, other estimators are more appropriate (see Besag [4] for an insightful discussion). We have found, for example, that the posterior mean ($\int x\Pi(dx|y)$) is more effective for tomography, at least in our experiments (see §5).

The computational issues for posterior mean and MAP estimators are similar. For illustration we concentrate here on MAP estimation. In most applications we cannot hope to identify the true maximum a posteriori image vector x. To appreciate the computational difficulty, consider again the Ising model with added white noise:

$$\Pi(x|y) = \frac{1}{z_p}\exp\{\beta\sum_{[s,t]} x_s x_t - \frac{1}{2\sigma^2}\sum_{s\in S}(y_s - x_s)^2\} \tag{2.3}$$

This is to be maximized over all possible vectors $x = \{x_s\}_{s\in S} \in \{-1,1\}^S$. With $|S| \sim 10^5$, brute force approaches are intractable; instead, we will employ a Monte Carlo algorithm which gives adequate approximations. (Remarkably, this particular example can be solved exactly, using techniques from network flow theory, see [29].)

Maximizing (2.3) amounts to minimizing

$$U_p(x) = -\beta\sum_{[s,t]} x_s x_t + \frac{1}{2\sigma^2}\sum_{s\in S}(y_s - x_s)^2$$

which might be thought of as the posterior energy. (As with z_p, the fixed observation y is suppressed in the notation $U_p(x)$.) More generally, we write the posterior distribution as

$$\frac{1}{z_p}\exp\{-U_p(x)\} \tag{2.4}$$

and characterize the MAP estimator as the solution to the problem

$$\text{choose x to minimize } U_p(x)$$

The utility of this point of view is that it suggests a further analogy to statistical mechanics, and a computation scheme for approximating the MAP estimate, which we shall now describe.

COMPUTING. When exploring a specific application, it is our consistent experience that computation-intensive algorithms for image analysis can be made fast by suitable compromises and exploitation of special structure. Still, as a research tool, it has been invaluable to have available a general computational framework, which we now describe in the context of MAP estimation.

Pretend that (2.4) is the equilibrium Gibbs distribution of a real system. Recall that MAP estimation amounts to finding a minimal energy state. For many physical systems the low-energy states are the most ordered, and these often have desirable properties. The state of silicon suitable for wafer manufacturing, for example, is a low-energy state. Physical chemists achieve low-energy states by heating and then slowly cooling a substance. This procedure is called *annealing*. Cerný [10] and Kirkpatrick [40] suggest searching for good minimizers of $U(\cdot)$ by *simulating* the dynamics of annealing, with U playing the role of energy for an (imagined) physical system. In our image processing experiments, we often use simulated annealing to find an approximation to the MAP estimator.

Dynamics are simulated by producing a Markov chain, $X(1), X(2), \ldots$ with transition probabilities chosen so that the equilibrium distribution is the posterior (Gibbs) distribution (2.4). One way to do this is with the Metropolis algorithm [47]. More convenient for image processing is a variation we call *stochastic relaxation*. The full story can be found in [22] and [33]. Briefly, in stochastic relaxation we choose a sequence of sites $s(1), s(2), \ldots \in S$ such that each site in S is visited infinitely often. If $X(t) = x$, say, then $X_r(t+1) = x_r$, $\forall r \neq s(t), r \in S$, and $X_{s(t)}(t+1)$ is a sample from

$$\Pi(X_{s(t)} = \cdot | X_r = x_r, r \neq s(t)),$$

the conditional distribution on $X_{s(t)}$, given $X_r = x_r$, $\forall r \neq s(t)$. By the Markov property,

$$\Pi(X_{s(t)} = \cdot | X_r = x_r, r \neq s(t)) = \Pi(X_{s(t)} = \cdot | X_r = x_r, r \in G^p_{s(t)})$$

where $\{G^p_s\}_{s \in S}$ is the *posterior* neighborhood system, determined by the posterior energy $U_p(\cdot)$. The prior distributions that we have experimented with have mostly had local neighborhood systems, and usually the posterior neighborhood system is also more or less local as well. This means that $|G^p_{s(t)}|$ is small, and this makes it relatively easy to generate, Monte Carlo, $X(t+1)$ from $X(t)$. In fact, if Ω is the range of $X_{s(t)}$, then

$$\Pi(X_{s(t)} = \alpha | X_r = x_r, r \in G^p_{s(t)}) = \frac{\Pi(\alpha, {}_{s(t)}x)}{\sum_{\hat{\alpha} \in \Omega} \Pi(\hat{\alpha}, {}_{s(t)}x)} \tag{2.5}$$

where

$$(\alpha, {}_{s(t)}x)_r = \begin{cases} \alpha & r = s(t) \\ x_r & r \neq s(t) \end{cases}$$

Notice that (fortunately!) there is no need to compute the posterior partition function z_p. Also, the expression on the right-hand side of (2.5) involves only those potential terms associated with cliques containing $s(t)$, since all other terms are the same in the numerator and the denominator.

To simulate annealing, we introduce an artificial temperature into the posterior distribution:

$$\Pi_T(x) = \frac{\exp\{-U_p(x)/T\}}{Z_p(T)}$$

As $T \to 0, \Pi_T(\cdot)$ concentrates on low energy states of U_p. To actually find these states, we run the stochastic relaxation algorithm while slowly lowering the temperature. Thus $T = T(t)$, and $T(t) \downarrow 0$. $\Pi_{T(t)}(\cdot)$ replaces $\Pi(\cdot)$ in computing the transition $X(t) \to X(t+1)$. In [22] we showed that, under suitable hypotheses on the sequence of site visits, $s(1), s(2), \ldots$:

If $T(t) > c/(1 + \log(1 + t))$, and $T(t) \downarrow 0$, then for all c sufficiently large $X(t)$ converges weakly to the distribution concentrating uniformly on $\{x : U(x) = min_y U(y)\}$.

More recently, our theorem has been improved upon by many authors. In particular, the smallest constant c which guarantees convergence of the annealing algorithm to a global minimum can be specified in terms of the energy function U_p (see [27] and [37]). Also, see Gidas [28] for some ideas about faster annealing via renormalization group methods.

In the experiments to be described here, MAP estimates are approximated by using the annealing algorithm or deterministic variations of annealing. This involves Monte Carlo computer generation of the sequence $X(1), X(2), \ldots$, terminating when the state ceases to change substantially.

3. TEXTURE SEGMENTATION AND CLASSIFICATION

Texture *synthesis* refers to computer generation of homogeneous patterns, usually intended to match a natural texture such as wood, grass, or sand. In many instances, Markov random fields provide good models, and Metropolis-like Monte Carlo methods yield respectable facsimiles of the real textures ([15],[17]). Here we combine MRF texture models, for the pixel process, with an Ising-like texture-label process, in order to segment and label a scene consisting of patches of natural textures. The image model thereby involves both a pixel process, of grey-level intensities, and a label process, whose components identify the texture type of each picture element in the scene. Our approach is similar to those of Derin and Elliott [17] and Cohen and Cooper [14], especially in our use of this two-tiered image model.

IMAGE MODEL. The image process comprises a pixel process and a label process, $X = \{X^P, X^L\}$. As usual, the pixel sites form an $N \times N$ square lattice, say S^P. For each pixel site there is a corresponding label site, and thus the graph associated with the image model has sites $S = S^P \cup S^L$, where S^L is just a copy of S^P. The elements of S^P and S^L index the components of X^P and X^L, respectively, so that $X^P = \{X^P_s\}_{s \in S^P}$ and

$X^L = \{X_s^L\}_{s \in S^L}$. In the experiments reported here, the pixels were allowed sixteen possible grey levels $X_s^P \in \{0,1,...,15\}, \forall s \in S^P$, whereas the range of the labels depended on the actual number of textures in the scene, thus assuming this number to be known a priori. Let M be the number of textures that are to be modelled. Then $X_s^L \in \{1,2,...,M\}, \forall s \in S^L$.

We shall develop the image model by first assuming that the texture type is fixed, say type l and constant over the scene. *Conditioned* on $X_s^L = l \in \{1,2,...,M\}, \forall s \in S^L$, the process X^P is a Markov random field:

$$\Pi(x^P | X_s^L = l, s \in S^L) = \frac{1}{z^{(l)}} \exp\{-U^{(l)}(x^P)\}$$

where $z^{(l)}$ is the usual normalizing constant

$$z^{(l)} = \sum_{x^P} \exp\{-U^{(l)}(x^P)\}$$

Only pair-cliques appear in the energy $U^{(l)}$. There are six types of pair-cliques, as shown in Figure 1. These we index by $i \in \{1,2,3,4,5,6\}$. We denote by $< s,t >_i$ a pair of sites s,t which form a type i clique, and by $\sum_{<s,t>_i}$ the summation over all such pairs. With these conventions, the (conditional) energy is

$$(3.1) \qquad U^{(l)}(x^P) = -\sum_{i=1}^{6} \sum_{<s,t>_i} \theta_i^{(l)} \Phi(x_s^P - x_t^P), \quad \Phi(\Delta) \doteq \left(1 + (|\Delta|/\delta)^2\right)^{-1}$$

for some fixed $\delta > 0$. Notice that $\Phi(x_s^P - x_t^P)$ is larger when $x_s^P = x_t^P$, and is monotonic in $|x_s^P - x_t^P|$. Because of this, the texture-dependent parameters $\theta_1^{(l)},...,\theta_6^{(l)}$ determine the degree to which neighboring pixels, of a particular type of pair-clique, will tend to have similar grey levels. In fact, if $\theta_i^{(l)} > 0$, then for texture l we expect pixel pairs x_s and x_t, of clique type i, to typically have similar intensities. If $\theta_i^{(l)} < 0$ then the tendency is to be different. Of course, these simple rules are complicated by the actions of the other five types of pair-cliques.

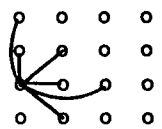


Figure 1.

The parameters $\theta_i^{(l)}$, for $i = 1,2,...,6$, and $l = 1,2,...,M$ are estimated from pictures of the M textures, as explained in the subsection on Parameter Estimation, below. On the

other hand, Φ, and indeed the neighborhood structure, is ad hoc. We have used Φ extensively in other applications in which our main concern is with the difference of intensities between neighboring pixels. Of course the quadratic, $\Phi(\Delta) = \Delta^2$, is simpler, but it unduly penalizes large differences.

Having modeled the M textures, we now construct a composite Markov random field which accounts for both texture labels, $X^L = \{X_s^L,\ s \in S^L\}$, and grey levels, $X^P = \{X_s^P,\ s \in S^P\}$. The joint distribution is

$$\Pi(X^P = x^P, X^L = x^L) = \frac{\exp\{-U_1(x^P, x^L) - U_2(x^L)\}}{z} \tag{3.2}$$

in which U_2 promotes label bonding (we expect the textures to appear in patches rather than interspersed) and U_1 specifies the interaction between labels and intensities. Specifically, we employ a simple Ising-type potential for the labels:

$$U_2(x^L) = -\beta \sum_{[s,t]} 1_{x_s^L = x_t^L} + \sum_{s \in S} w(x_s^L), \quad \beta > 0 \tag{3.3}$$

Here β determines the degree of clustering, $[s,t]$ indicates a pair of nearest horizontal or vertical neighbors, and $w(\cdot)$ is adjusted to eliminate bias in the label probabilities (more on the choice of $w(\cdot)$ later).

To describe the interaction between labels and pixels we introduce the symbols $\tau_1, \tau_2, \ldots, \tau_6$ to represent the lattice vectors associated with the 6 pair-cliques (Figure 1). Thus s and $s + \tau_i$ are neighbors, constituting a pair with clique type i. The interaction is then given in terms of pixel-based contributions,

$$H(x^P, l, s) \doteq -\sum_{i=1}^{6} \theta_i^{(l)} \left\{\Phi(x_s^P - x_{s+\tau_i}^P) + \Phi(x_s^P - x_{s-\tau_i}^P)\right\} \tag{3.4}$$

and local sums of these called block-based contributions,

$$Z(x^P, l, s) \doteq \frac{1}{a} \sum_{t \in N_s} H(x^P, l, t). \tag{3.5}$$

Here, N_s is a block of sites centered at s (5×5 in all of our experiments), and the constant a is adjusted so that the sum of all block-based contributions reduces to $U^{(l)}$ (see(3.1)):

$$U^{(l)}(x^P) = \sum_{s \in S} Z(x^P, l, s) \tag{3.6}$$

This amounts to ensuring that each pair-clique appears exactly once ($a = 50$, for example, when N_s is 5 by 5). In terms of (3.4) and (3.5), the interaction energy, $U_1(x^P, x^L)$, is written

$$U_1(x^P, x^L) = \sum_{s \in S} Z(x^P, x_s^L, s). \tag{3.7}$$

Because of (3.6), the model is consistent with (3.1) for homogeneous textures, $X_s^L = l$, $\forall s \in S$. The idea is that each local texture label, X_s^L, is influenced by the pixel grey levels in a neighborhood of s.

Finally, to clarify the bias correction term $w(\cdot)$, we briefly examine the local characteristics of the field, specifically the conditional distributions for the labels given all the intensity data and the values of the neighboring labels. (The actual neighborhoods of the Markov random field corresponding to (3.2) can be easily inferred from (3.3) and (3.7).) The log odds of texture type k to type j is

$$\begin{aligned}
&\log\left\{\frac{\Pi(X_r^L = k | X_s^L = x_s^L,\ s \neq r\ ;\ X_s^P = x_s^P,\ s \in S)}{\Pi(X_r^L = j | X_s^L = x_s^L,\ s \neq r\ ;\ X_s^P = x_s^P,\ s \in S)}\right\} \\
&\quad = Z(x^P, j, r) - Z(x^P, k, r) + \beta \sum_{t:[t,r]} (1_{x_t^L = k} - 1_{x_t^L = j}) + \omega(j) - \omega(k) \\
&\quad = \frac{1}{z} \sum_{i=1}^{6} \sum_{s \in N_r} (\theta_i^{(k)} - \theta_i^{(j)}) \left\{\Phi(x_s^P - x_{s+\tau_i}^P) + \Phi(x_s^P - x_{s-\tau_i}^P)\right\} \\
&\qquad + \beta \sum_{t:[t,r]} (1_{x_t^L = k} - 1_{x_t^L = j}) + \omega(j) - \omega(k)
\end{aligned}$$

The first term imposes fidelity to the data x^P, and the second bonds the labels. The efficacy of the model depends on the extent to which the first term separates the two types k and j, which can be assessed by plotting histograms for the values of this quantity both for pure k and pure j data. A clean separation of the histograms signifies a good discriminator. However, since we are looking at log odds, we insist that the histograms straddle the origin, with positive (resp. negative) values associated with texture type k (resp. j). The function $w(\cdot)$ makes this adjustment.

DEGRADATION MODEL. The degradation is deterministic. The observation process is the pixel process $Y = X^P$, and hence the degradation is just the *projection* $(X^P, X^L) \to X^P$.

POSTERIOR DISTRIBUTION. In this special case, the posterior energy is the same as the prior energy, but some of the components are fixed. In particular

$$\Pi\left((x^P, x^L)|y\right) = \frac{1}{z_p} \exp\left\{-U_1(x^P, x^L) - U_2(x^L)\right\} 1_{x^P = y}$$

Equivalently, we simply use $\Pi(x^L|x^P)$ as the posterior distribution:

$$\Pi(x^L|x^P) = \frac{1}{z_p} \exp\left\{-U_1(x^P, x^L) - U_2(x^L)\right\}$$

ESTIMATING THE IMAGE. We use MAP estimation: Given an observation, $X^P = x^P$, we seek x^L to minimize $U_1(x^P, x^L) + U_2(x^L)$.

COMPUTING. We use stochastic relaxation, with simulated annealing, as described in §2. A convenient starting point is arrived at by "turning off" the Ising term in the label model (3.3): we set $\beta = 0$. Since this is the only label/label interaction term in the model, the MAP estimate of x^L, with $\beta = 0$, is determined by (locally) optimizing x_s^L at each $s \in S^L$. The computation time is negligible. Thereafter, we set β to the model value (see the subsection on Parameter Estimation, below) and begin stochastic relaxation. In the experiments, each site was visited about 150 times.

EXPERIMENTAL RESULTS. Three experiments were done on texture discrimination, based on two images with two textures each and one with four. There are four textures involved: wood, plastic, carpet, and cloth. As mentioned above, the parameters were estimated from the pure types (see the subsection on Parameter Estimation, below). There was no pre- or post-processing. In particular, no effort was made to "clean-up" the boundaries, expecting smooth transitions. The results are shown in Figures 2, 3, and 4; these correspond to i) wood on plastic, ii) carpet on plastic, and iii) wood, carpet, and cloth on plastic background.

FIGURES 2,3,4 ARE PLACED AT THE END OF THE TEXT.

In each figure, the left panel is the textured scene, and the right panel shows the segmentation, with texture labels coded by grey level. It is interesting to note that the grey level histograms of the four textures are very similar (Figure 5); in particular, discrimination based on shading alone is virtually impossible.

The model is not really adequate for texture *synthesis*; samples generated from the model do not resemble the texture very well. Evidently, the utility of Markov random field models does not depend on their capacity for simulating real-world imagery. A more serious drawback of our model is that it is dedicated to a fixed repertoire of textures, viewed at a particular orientation and at a particular magnification, or range. The problem is easier if the goal is merely *segmentation*, without *recognition*. We are experimenting with segmentation algorithms that are basically scale and orientation independent. Indeed, there are no texture-specific parameters. These are built upon the same modelling/computing

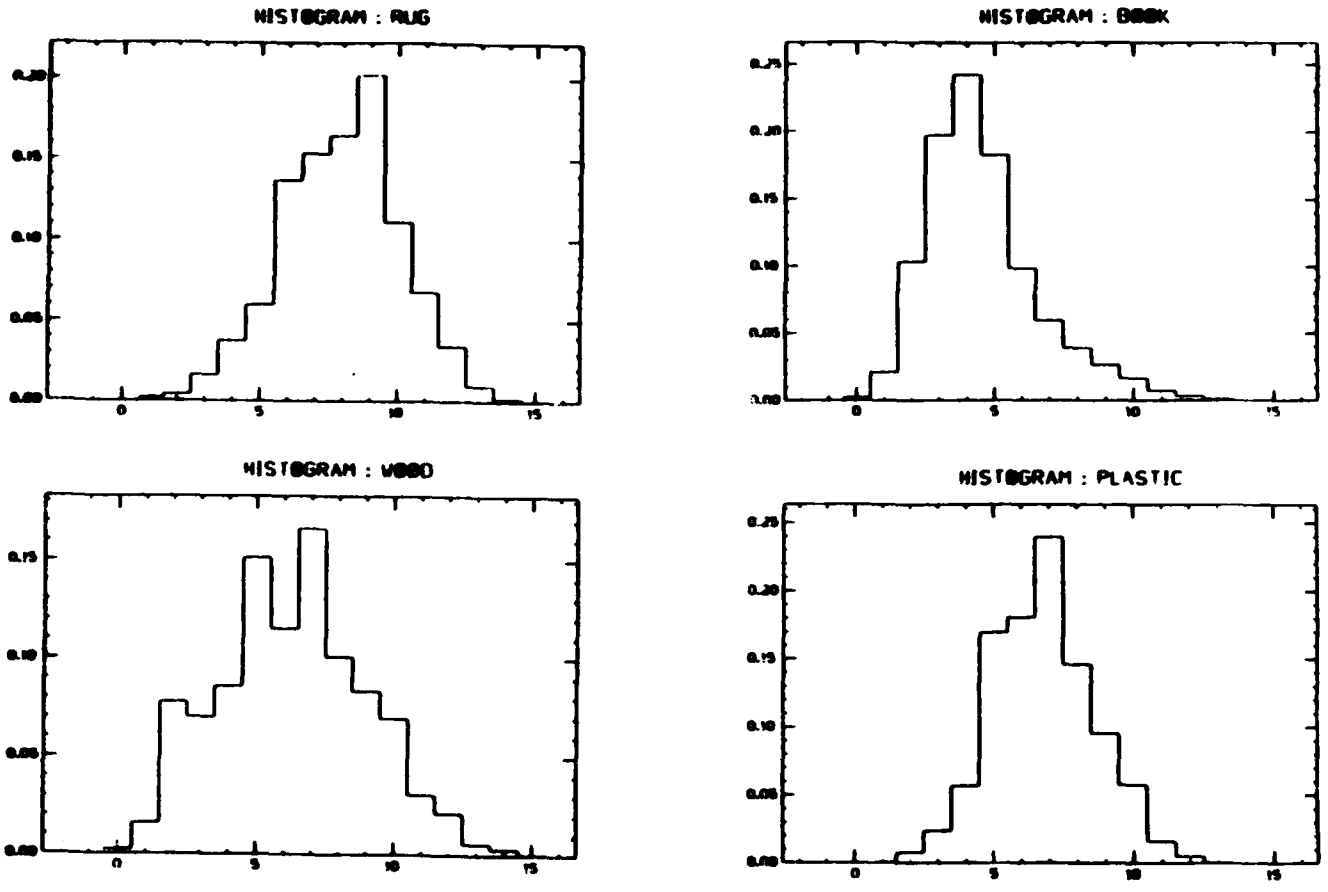


Figure 5. Grey Level Histograms.

framework.

PARAMETER ESTIMATION. The performance of the model is not unduly sensitive to the choice of δ (see (3.1)) or β (see (3.3)), which were determined by trial and error. On the other hand, the pair-clique parameters $\theta_i^{(l)}, i = 1,2...,6, l = 1,2,...,M$ characterize the M textures, and critically determine the ability of the model to segment and label. Needless to say, these must be systematically estimated. Trial and error is not feasible.

We have estimated the parameters from samples of the M textures. These training samples contain only one texture each, and we used just one sample for each texture. For a fixed texture, say wood, and from a single sample, say $\tilde{x}^P$, the problem then is to estimate $\theta_1, \theta_2, ..., \theta_\ell$ in the model

$$\Pi(X^P = x^P; \theta) = \frac{\exp\{-U(x^P; \theta)\}}{z(\theta)}$$

where

$$U(x^P;\theta) = -\sum_{i=1}^{6}\sum_{<s,t>_i}\theta_i\Phi(x_s^P - x_t^P)$$

and

$$z(\theta) = \sum_{x^P}\exp\{-U(x^P;\theta)\}$$

(We include $\theta \doteq (\theta_1,...,\theta_6)$ in Π, U, and Z to emphasize the dependencies on the unknown parameters.) The standard approach is to maximize the likelihood:

$$\text{choose } \theta \text{ to maximize } \Pi(\tilde{x}^P;\theta)$$

Of course, maximizing Π is equivalent to maximizing $\log\Pi$. It is easily demonstrated that the latter is *concave* in θ with gradient

$$(3.8)\qquad \nabla \log\,\Pi(\tilde{x}^P;\theta) = \left\{\sum_{<s,t>_i}\Phi(\tilde{x}_s^P - \tilde{x}_t^P) - E_\theta\left[\sum_{<s,t>_i}\Phi(X_s^P - X_t^P)\right]\right\}_{i=1,...,6}$$

where $E_\theta[\cdot]$ is expectation with respect to $\Pi(\cdot;\theta)$. This suggests a gradient ascent procedure, but the expectation $E_\theta[\cdot]$ is intractable, involving summation over the entire range of X^P. In our experiments, we used a 16 grey-level scale for the pixels, and 204×204 lattices; then the expectation in (3.8) has $16^{(204\times204)}$ terms. An alternative to brute force evaluation is to use stochastic relaxation (see §2), which produces an (asymptotically) *ergodic* sequence $X^P(1), X^P(2),...$ for any given θ, and from which expectations can be approximated by appropriate time-averages. This, too, is computationally intensive, but feasible. In some settings we have found no alternative, and this Monte Carlo procedure has worked well, albeit slowly (see [43]). See also Hinton and Sejnowski [39] for a closely related algorithm, used to model learning in a theory of neuron dynamics.

For homogeneous random fields, such as our image models, Besag ([2],[3]) has proposed an ingenious alternative to maximum likelihood, known as maximum *pseudolikelihood*. The pseudolikelihood function is

$$PL(\tilde{x}^P;\theta) \doteq \prod_{s\in S^P\setminus\partial S^P}\Pi(X_s^P = \tilde{x}_s^P | X_r^P = \tilde{x}_r^P, r \neq s;\theta)$$

where ∂S^P is the boundary of S^P under the neighborhood system determined by the energy U, and $S^P\setminus\partial S^P$ is the complement of ∂S^P relative to S^P. The *pseudolikelihood estimator* is the θ that maximizes $PL(\tilde{x}^P;\theta)$. Recently, we have lent some analytic support by establishing consistency of pseudolikelihood in the large-graph limit (see Geman and Graffigne

[23]). The result is independent of *critical phenomena*, or *lack of spatial stationarity*, both of which can occur with infinite volume Gibbs states having translation-invariant potentials [54].

We emphasize the overwhelming computational advantage of pseudolikelihood. As with the log likelihood function, the log pseudolikelihood function, $\log PL(\tilde{x}^P;\theta)$, is concave, but this time the gradient is directly computable:

$$\nabla \log\ PL(\tilde{x}^P;\theta) = \nabla \sum_{s\in S^P\setminus\partial S^P} \left\{ \sum_{i=1}^{6} \theta_i \left\{ \Phi(\tilde{x}_s{}^P - \tilde{x}^P_{s+\tau_i}) + \Phi(\tilde{x}^P_s - \tilde{x}^P_{s-\tau_i}) \right\} \right.$$

$$\left. - \log \sum_{\alpha} \exp \left\{ \sum_{i=1}^{6} \theta_i \left\{ \Phi(\alpha - \tilde{x}^P_{s+\tau_i}) + \Phi(\alpha - \tilde{x}^P_{s-\tau_i}) \right\} \right\} \right\}$$

(where $\sum_\alpha$ is summation over pixel grey levels, zero through fifteen in our experiments)

$$= \left\{ \sum_{s\in S^P\setminus\partial S^P} \left\{ \Phi(\tilde{x}^P_s - \tilde{x}^P_{s+\tau_i}) + \Phi(\tilde{x}^P_s - \tilde{x}^P_{s-\tau_i}) \right.\right.$$

$$\left.\left. - E_\theta \left[\Phi(X^P_s - X^P_{s+\tau_i}) + \Phi(X^P_s - X^P_{s-\tau_i}) | X^P_r = \tilde{x}^P_r, r \neq s \right] \right\} \right\}_{i=1,\ldots,6}$$

This time, the expectation is tractable. The conditional distribution on X^P_s, given $X^P_r = \tilde{x}^P_r, r \neq s$, involves only those variables $\tilde{x}^P_r$ in the neighborhood of s. Furthermore, this time the summation for the expected value is over the range of X^P_s only, which has only sixteen values. In short, the gradient of the log pseudolikelihood is directly computable, and therefore gradient ascent is feasible without resorting to time-consuming Monte Carlo methods. For the texture experiments discussed above, the pair-clique parameters were estimated, for each texture type, by gradient ascent of the pseudolikelihood function.

Some modifications of maximum likelihood and pseudolikelihood have been recently introduced by Chalmond [11]. A third alternative was suggested by Derin and Elliott ([17],[18]), and has been studied and analyzed extensively by Possolo [53]. This involves a regression fit of the log of the local conditional probabilities, and works best when there are a small number of values in the range of the random variables. For example, the method is very effective for Ising-like models.

DIRECTIONS FOR FUTURE RESEARCH. We are continuing to study the following related issues:

(1) Performance of Estimators. Rates of convergence, and error bounds, for parameters estimated by pseudolikelihood merit further study. We hope to deepen our understanding

of how pseudolikelihood estimators compare to maximum likelihood estimators and to work in the direction of an elementary distribution theory for the pseudolikelihood method.

(2) Better Models for Texture Synthesis. We shall begin with the recent advances by Cohen and Cooper [14].

(3) Model-independent Segmentation. We have begun development of a method for *segmenting* textured images without classifying the types of textures present. This approach requires no training. Preliminary versions of the model have performed well, but our analytic understanding is limited. The prior is again a Markov random field, but on a *random graph structure*. The random graph is employed to simplify computation of the MAP estimator. Indeed, experiments suggest that the posterior distribution has *no local minima*, and hence lends itself to simple iterative improvement. We will try to settle the conjecture that, in a large graph limit, there are no local minima in the energies of this class of Markov random fields with random graphs.

The classical Ising model provides a simple example. The usual, nearest-neighbor, energy is

$$U(x) = -\sum_{[s,t]} x_s x_t$$

where $x_s \in \{-1, 1\}, s \in S_N \doteq N \times N$ square lattice, and $\sum_{[s,t]}$ is summation over all pairs of horizontal and vertical nearest neighbors. Of course the two configurations that achieve the minimum of U are obvious (all +1's and all -1's), but nevertheless difficult to find by the usual Metropolis-like relaxation methods. This is because there are many flat ridges and local minima in the energy landscape. Consider, instead, a *random graph* S_N, with fixed degree (say 4), and let this define the neighborhood system in the expression for U. Our simulations suggest that for N sufficiently large U has no ridges or local minima; it can be minimized by iterating through S_N, sequentially changing the states x_s to create the lowest contribution to U. In particular, when N is sufficiently large there will be no "ties": *every* non-minimal configuration has at least one site s such that changing the sign of x_s strictly lowers the energy. Random graphs have been key in making our model-independent segmentation algorithm computationally feasible.

4. BOUNDARY DETECTION

Here we study boundary placement (without segmentation per se) including the problems of locating sudden changes in depth (occluding boundaries) or shape (surface creases, etc.). The idea is to define contours which are faithful to the 3-D scene but to avoid the non-physical edges due to noise, digitization, texture, lighting, etc. Obviously, there are global discontinuities, such as shadows, which are essentially impossible to distinguish from the occluding and shape boundaries, at least without higher level information or information from multiple sensors, in which case boundary *classification* becomes possible.

The complications are well-known: digital edges tend to be very noisy, due in part to the digitization process itself, but also to de-focusing and random effects in detecting the photons. The result is a variety of pathologies: "true" boundaries suddenly disappear, spurious ones appear haphazardly, and in general the surface transitions are highly redundant.

We formulate boundary detection as a single optimization problem, fusing the detection of edges, as well as pruning, linking, smoothing, etc.; there is no surface reconstruction. The subject of edge detection is wildly active, and there has been considerable progress of late in designing filters based on differential operators for optimally detecting various ideal step, crease, and other edges in noise-corrupted 1-D and 2-D signals ([9],[44],[56]). Other methods detect edges after fitting smooth surfaces to the data ([38],[36],[50]). Still others ([5],[49],[56],[45]) perform surface reconstruction and boundary detection at the same time, and are cast in a framework similar to the set-up in [22].

The use of boundary maps as the input to further processing is ubiquitous in computer vision; for example, algorithms for stereopsis, optical flow, and simple object recognition are often based on matching boundary segments. Other applications include the analysis of medical images (e.g. angiograms and ultrasound); automated navigation [6]; and the detection of the paths of roads and geologic faults, or the edges of lakes, flood plains, and crop fields, in remotely-sensed images.

IMAGE MODEL. We will give an outline. The details can be found in [21].

Denote the pixel lattice $\{(i,j) : 1 \leq i,j \leq N\}$ by S^P and let S^B denote another regular lattice interspersed among the pixels (see Figure 6(a)) and of dimension $(N-1) \times (N-1)$; these are the *boundary sites*.

We will mimic the pixel/label formulation used in §3: $X = \{X^P, X^B\}$, a pixel process and a boundary process, with $X^P = \{X^P_s\}_{s \in S^P}$ and $X^B = \{X^B_s\}_{s \in S^B}$. The observation

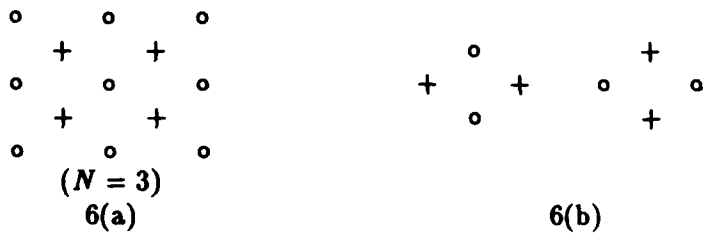


Figure 6.

process, or data, consists of grey-levels $Y_s, s \in S^P$, possibly distorted due to the imaging and recording system. Again we take $X \rightarrow Y$ to be a projection, so that $Y = X^P$, assuming undegraded observations of the pixel intensities. (See [22] for boundary finding *with* restoration.) Given $y = x^P$, we wish to assign values to the boundary variables $x^B = \{x_s^B, s \in S^B\}$, where $x_s^B = 1$ (resp. 0) indicates the presence (resp. absence) of a boundary at site $s \in S^B$. We have already discussed the corresponding interpretation of boundary map $x^B = x^B(y)$ in terms of physical discontinuities in the underlying three-dimensional scene.

We establish a boundary resolution or grid size $\sigma \geq 1$. (This is distinct from the resolution of the pixels, although the two are naturally related.) Let $S^B_{(\sigma)} \subset S^B$ denote the sub-lattice $\{(i\sigma+1, j\sigma+1) : 0 \leq i,j \leq (n-2)/\sigma\}$. Only the variables x_s^B, $s \in S^B_{(\sigma)}$, interact directly with the data; the remaining variables x_s^B, $s \in S^B \setminus S^B_{(\sigma)}$, are determined by those on the grid $S^B_{(\sigma)}$. Figure 7 shows the grids $S^B_{(2)}$ and $S^B_{(3)}$ for N=8; the sites off the grid are denoted by dots. Let Ω^P and Ω^B denote the state spaces of intensity arrays and boundary maps respectively, then

$$\Omega^P = \{\{x_s^P\} : s \in S^P, 0 \leq x_s^P \leq 255\}, \text{ and } \Omega^B_{(\sigma)} = \{\{x_s^B\} : s \in S^B_{(\sigma)}, x_s^B \in \{0,1\}\}.$$

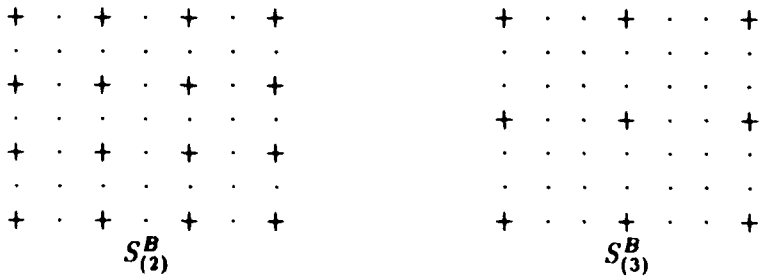


Figure 7.

The selection of σ influences the interpretation of x^B, the computational load, the interaction range at the pixel level, and is related to the role played by the size of the spatial filter in edge detection methods based on differential operators.

Let $< s,t >_\sigma$, $s,t \in S^B_{(\sigma)}$ denote a nearest-neighbor pair relative to the grid. Thus, $s = (i\sigma + 1, j\sigma + 1)$, $t = (k\sigma + 1, l\sigma + 1)$ is such a (horizontal or vertical) pair if either $i = k$ and $j = l \pm 1$, or $j = l$ and $i = k \pm 1$. We identify $< s,t >_\sigma$ with the elementary boundary segment consisting of the horizontal or vertical string of $\sigma + 1$ sites (in S^B) including s,t and the $\sigma - 1$ sites between s and t.

A *disparity measure* $\Delta_{st} = \Delta^{(\sigma)}_{st}(x^P)$ will gauge the intensity *flux* across $< s,t >_\sigma$, i.e. orthogonal to the associated segment. We will experiment with several types of measures. An obvious choice at high resolution ($\sigma = 1$) is $\Delta_{st} = |x^P_{s^*} - x^P_{t^*}|$ (or a suitable linear transformation of this) where $s^*, t^* \in S^P$ are the two pixels associated with $< s,t >_\sigma$; see Figure 6(b). The analogous choice at a lower resolution ($\sigma > 1$) might be a measure of the form $\Delta_{st} = m^{-1}|\sum_{D_1} x^P_t - \sum_{D_2} x^P_t|$, where $D_1, D_2 \subset S^P$ are *adjacent blocks of pixels*, of the same size (m) and shape, and separated by $< s,t >_\sigma$. These and other measures are discussed later on.

The energy function $U(x^P, x^B)$ should promote boundary maps x^B which are faithful to the data $y = x^P$ in the sense that large values of $\Delta_{st}(x^P)$ are associated with "on" segments ($x^B_s x^B_t = 1$) and small values with "off" segments ($x^B_s x^B_t = 0$). There are no *a priori* constraints on x^B at this point; in fact, for most actual scenes x^P, the energy U will actually favor maps x^B with undesirable dead-ends, multiple representations, high curvature, etc. A simple choice for the x^P/x^B interaction is

$$U(x^P, x^B) = \sum_{<s,t>_\sigma} (1 - x^B_s x^B_t)\Delta_{st}(x^P)$$

where the summation extends over all nearest-neighbor pairs $< s,t >_\sigma$.

Returning to the disparity measure, we employ one type for depth and shape boundaries and another for the texture experiments. In the former case, the disparity measure involves the (raw) grey-levels only, whereas for texture discrimination we also consider data transforms based on so-called directional residuals. Except when $\sigma = 1$, the data sets are compared by the Kolmogorov-Smirnov distance [7], as explained below.

At the highest resolution ($\sigma = 1$), the simple measure $|x^P_{s^*} - x^P_{t^*}|$ suitably normalized (where s^*, t^* are the two pixels associated with the boundary sites s,t; see Figure 6(b)) can be effective for simple scenes.

At lower resolution, let D_1, D_2 denote two adjacent blocks of pixels, of equal size and shape. An example is illustrated in Figure 8 for the case of two square blocks of size $5^2 = 25$ which straddle a vertical segment $< s,t >_\sigma$ with $\sigma = 3$. Let $x^P(D_j) = \{x^P_s, s \in D_j\}$,

$j = 1, 2$, be the grey-levels in the blocks and set

$$\Delta_{s,t}(x^P) = d\left(x^P(D_1), x^P(D_2)\right)$$

where $d(\cdot,\cdot)$ is a monotonic function of the Kolmogorov-Smirnov distance between two empirical distributions. This, for example, is the type of disparity measure we use for the house scene and the radar scene (shown below).

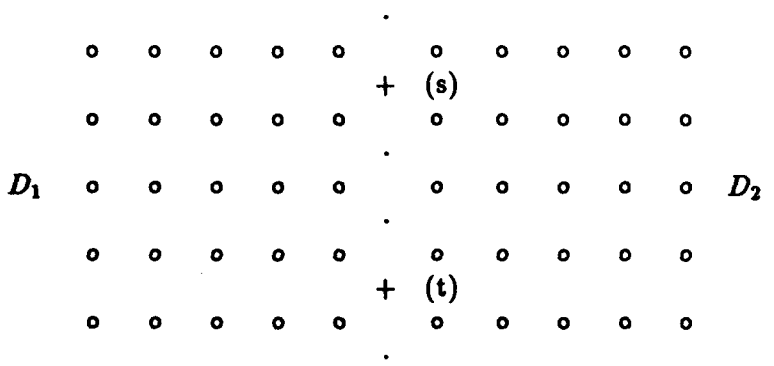


Figure 8.

Various enhancements in Δ are necessary to detect the more subtle transitions sometimes found between two textured regions. These include the use of multiple, high-order, spatial features in the definition of d. In the experiments below, several such features were used in segmenting the Brodatz texture collage.

The other component in the model is a *penalty function* $V(x^B)$ which counts the number of "taboo" patterns in x^B; states x^B for which $V(x^B) > 0$ are "forbidden." For example, boundary maps are penalized for dead-ends, clutter, density, etc. $V(x^B)$ denotes the total number of "penalties" associated with $x^B \in \Omega^B_{(\sigma)}$. These penalties are simply local binary patterns over subsets of $S^B_{(\sigma)}$. Figure 9 illustrates a family of four such patterns; they can be associated with any resolution σ by the obvious scaling.

Figure 9.

The patterns in Figure 9 correspond, respectively, to an isolated or abandoned segment, sharp turn, quadruple junction, and too-small structure. Depending on σ, the pixel resolution, and scene information, we may or may not wish to include the latter three. For

example, including the last one with $\sigma = 6$ would prohibit detection of a square structure of pixel size 6×6.

Finally, our image model is

$$\Pi(x^P, x^B) = \frac{1}{z} e^{-U(x^P, x^B)} 1_{V(x^B)=0},$$

the usual Gibbs distribution, but confined to the allowed boundary states.

DEGRADATION MODEL. The degradation is again deterministic; we observe the projected data $y \doteq x^P \leftarrow (x^P, x^B)$.

POSTERIOR DISTRIBUTION. As in §3, this is just

$$\Pi\left((x^P, x^B)|y\right) = \frac{1}{z_p} \exp\left\{-U(x^P, x^B)\right\} 1_{V(x^B)=0} 1_{x^P=y}.$$

ESTIMATING THE IMAGE. Our experiments in boundary finding have been with (approximate) MAP estimators.

COMPUTING. This again is by stochastic relaxation. An extension of the theory outlined in §2 is needed to handle the *constrained* optimization problem

$$\text{minimize}_{x^B : V(x^B)=0} \sum_{<s,t>_e} (1 - x_s^B x_t^B) \Delta_{s,t}(x^P).$$

By introducing the constraints slowly during the annealing sequence, a global minimum, within the constrained subset, can again be found. The theory behind this enhancement of stochastic relaxation is layed out in [21].

EXPERIMENTAL RESULTS. In each of Figures 10, 11, and 12 the original scene is in the top panel. A version of stochastic relaxation is used to incorporate constraints while minimizing U. The placement of boundaries is shown in the bottom panel. Figure 10 is a synthetic aperture radar picture of water and ice (water is the region generally appearing darker), and Figure 11 is a collage of four Brodatz textures [8], chosen for the rather subtle transitions created by their juxtapositions. The bottom panel of Figure 11 shows the evolution of the boundary process during relaxation.

FIGURES 10,11,12, ARE PLACED AT THE END OF THE TEXT.

DIRECTIONS FOR FUTURE RESEARCH. These models may be extended in many directions. What follows is a brief description of two such generalizaitons.

(1) Other Boundary Primitives. The elementary boundary unit or primitive is a horizontal or vertical segment whose length is resolution-dependent, namely $\sigma + 1$ in pixel units. As a result, a discontinuity running at 45° is detected and localized with less reliability than one at 0° or 90°, where the disparity measure has maximum sensitivity. An obvious remedy is to replace the pairs $< s, t >_\sigma$ by other primitive segments, e.g. a distinguished family of triples, such as the six represented (up to translates) in Figure 13.

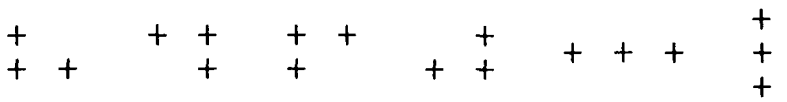


Figure 13.

More generally, for any such family $\mathcal{A}$ consider the interaction term

$$\sum_{A \in \mathcal{A}} (1 - \Pi_{s \in A} x_s^B) \Delta_A(x^P)$$

where $\Delta_A(x^P)$ is a measure of the disparity in the A-direction. In Figure 13, these directions are, respectively, $\pi/4$, $\pi/4$, $-\pi/4$, $-\pi/4$, $\pi/2$, and 0. One can imagine a variety of ways to situate two appropriately shaped sets of pixels to straddle (and abut) segments such as these.

(2) Multivariate Data. Data may be available from several sensors, e.g. multispectral satellite data, or an optical camera and laser radar. One might integrate these with a composite disparity measure of the type used in the texture studies, viz. $\Delta = \max \Delta^{(k)}$ in order to improve discrimination.

Another possibility is to attempt to classify boundaries, especially if range data is also available. Let $x^{P(1)}$ and $x^{P(2)}$ be range and brightness values, and let x_s^B assume three values, say 0,1,2, corresponding to "off", "occluding (=depth) boundary" and "other boundary" (for instance a crease or shadow). Now design the energy function to couple $\Delta(x^{P(1)})$ with $\xi = \{\xi_s\}$, $\xi_s = \delta_{\{1\}}(x_s^B)$ and $\Delta(x^{P(2)})$ with $\eta = \{\eta_s\}$, $\eta_s = \delta_{\{1,2\}}(x_s^B)$; for example, just add the two corresponding terms to form U. The penalty patterns are the usual ones regarding dead-ends, etc., as well as mixtures of 0,1,2 corresponding to physically implausible (or impossible) transitions.

5. SINGLE PHOTON EMISSION TOMOGRAPHY

Emission tomography is used to determine the distribution of a pharmaceutical in a part of the body such as the brain, liver, or heart. Depending upon the pharmaceutical used, this concentration can be taken as a measure of local blood flow (perfusion) and/or local metabolic activity. Glucose, for example, is taken up by neuronal cells in proportion to metabolic activity, and the latter generally mirrors recent electrical activity. Thus, areas of the brain most used in performing a cognitive or motor task will demonstrate a relatively increased uptake of glucose immediately following the task. For the heart, pharmaceuticals can be chosen whose uptake reflects local perfusion. The concentration of these pharmaceuticals can thereby be used to assess the adequacy of blood flow to the different parts of the heart.

In single photon emission computed tomography (SPECT), pharmaceutical concentration is estimated by detecting *photon emissions* from an injected or inhaled dose of the pharmaceutical that has been chemically combined with a radioactive isotope. This combined agent is called a radiopharmaceutical. The goal of SPECT is to determine radiopharmaceutical concentration (equivalently, isotope concentration or density) as a function of position in a region of the body. Detectors with collimators are strategically placed around the region of interest, and these are able to count photons emitted by radioactive decay of the isotope. A detector will capture those photons which escape attenuation and whose trajectories carry them down the bore of the collimator.

The determination from photon counts of isotope concentration as a function of position is referred to as *reconstruction*. Interest in *statistical* approaches to reconstruction problems in emission computed tomography was greatly enhanced by the work of Shepp and Vardi [55] on the use of maximum likelihood (ML) methods. There are earlier instances of suggestions to regard the reconstruction problem as a statistical estimation problem; however, the demonstration of the versatility of the approach as well as the specification of algorithms *that work* were advanced substantially by Shepp and Vardi's work.

The image reconstruction problem, viewed as an estimation problem, is inherently nonparametric: one seeks an estimate of a function of general form on a continuous domain. As such, it is widely recognized that the estimates need to be regularized or smoothed, especially in "small sample" implementations. Various approaches to regularization have been suggested, including penalized ML, the method of sieves, and Bayesian methods. In Geman and McClure [24], we proposed that *a priori* spatial information be built into a statistical reconstruction algorithm, in a Bayesian approach, by quantifying spatial constraints in the

form of a Gibbs prior distribution. In Geman and McClure [25] we demonstrated the feasibility of estimating parameters of the Gibbs prior from single patient data. A brief outline of the Bayesian approach, with parameter estimation, follows.

IMAGE MODEL. Let X_s denote the concentration of the radiopharmaceutical at the point $s = (x,y)$ in the domain Ω of interest. We shall take Ω to be a bounded two-dimensional region, though for the models and methods we will describe there are no essential changes when Ω is three-dimensional.

We shall construct a prior distribution on X that captures simple prior expectations about the qualitative nature of the isotope density. Mainly, we wish to exploit the anticipated smoothness of X. Neighboring locations will typically have similar intensity levels. But we must also accommodate sharp changes in concentration, which might occur across an arterial wall or across a boundary between two tissue types.

In the spirit of nonparametric estimation, we might construct the prior on a suitable space of functions $X : \Omega \rightarrow \mathrm{R}$. It is more convenient, however, to do the construction on the discrete lattice domain S introduced in §2. The prior, therefore, is on the array $X = \{X_s\}_{s \in S}$. As usual, we adopt the Gibbs representation (2.1).

The expected configurations are those for which typical neighboring sites $s, t \in S$ have similar intensities X_s, X_t. This is a *local* constraint and it is conveniently captured by a locally composed energy function U,

$$U(X) = \sum_{[s,t]} \beta\phi(X_s - X_t) + \sum_{<s,t>} \frac{\beta}{\sqrt{2}}\phi(X_s - X_t). \tag{5.1}$$

Here we use $[s,t]$ to indicate that s and t are nearest horizontal or vertical neighbors in the lattice S, and $< s,t >$ to denote diagonal neighbors. The constant β is positive and the function $\phi(\xi)$ is even and minimized at $\xi = 0$. Thus U is minimized by configurations of constant intensity. Under the Gibbs distribution with energy (5.1) the more likely isotope densities are those with small site-to-site variation in intensity.

To achieve the desired properties for the more likely isotope densities, the exact form of ϕ is probably not important, but its qualitative features can make a difference. We have experimented with ϕ's that are increasing in ξ for $\xi \geq 0$. An obvious choice is $\phi(\xi) = \xi^2$, but then under $\Pi(X)$, large intensity gradients, as would be associated with certain natural boundaries, are exceedingly unlikely. Instead, we use functions of the form

$$\phi(\xi) = \frac{-1}{1 + (\xi/\delta)^2} \tag{5.2}$$

where δ, like β, is a constant to be fixed later.

Levitan and Herman [41] have recently proposed the use of Gaussian priors in a Bayesian formulation. Liang and Hart [42] also suggest the use of Gaussian priors, as well as others, deduced by max-ent arguments from prior constraints on low-order moments of X. Our earlier experiments with the *quadratic* energy function indicated that the resulting Bayesian algorithms oversmoothed real boundaries where the difference $(X_s - X_t)$ should be allowed to be large. The finite asymptotic behavior of our ϕ-function was designed to mitigate this oversmoothing.

DEGRADATION MODEL. We assume that the detectors are arranged in a linear array, at equally spaced lateral sampling intervals, and that the detector array can be positioned at any orientation θ relative to the x-axis. (See Figure 14.) We assume the detectors are of so-called *parallel bore* type, meaning that they detect only those photons in a small interval $[\theta - \Delta\theta/2, \theta + \Delta\theta/2]$ when the array has orientation θ. Let L denote the total number of detectors in the array and let $\Delta\sigma$ denote the spacing between detectors.

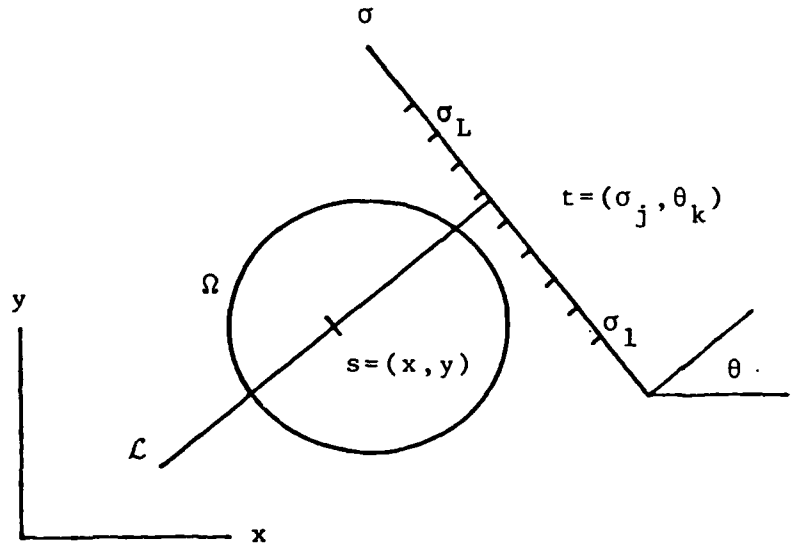


Figure 14. Detector Geometry.

The physical effects incorporated in the model are the *spatial Poisson process* that

describes the sites of the radioactive decays from which photons emanate and the *process of photon attenuation* by which photons are annihilated and their energy is absorbed by matter through which their trajectories pass. Attenuation is accurately described by a linear attenuation function $\mu(s)$ on Ω. The function μ is assumed to be known; values of μ for bone, muscle, etc. and for various photon energies are known *a priori* or could be measured by transmission tomographic methods. Attenuation is a memoryless process and we can thus deduce the functional form of the probability that a photon survives to reach the detector array. When a photon trajectory has direction θ and it emanates from site $s = (x, y)$ in Ω, then

$$P(\text{photon survival}) = \exp\{-\int_{\mathcal{L}(x,y)} \mu(\xi,\eta)dl\},$$

where the line integral is taken over the segment $\mathcal{L}(x, y)$ from (x, y) to the detector and dl is differential arc length.

For our sampling design, we shall position the detector array at n equally spaced angles θ_k for duration T time units at each angle. Then at each angle, we observe the random variables Y_t, for $t \in D_k = \{(\sigma_j, \theta_k), j = 1, \ldots, L\}$ that give the numbers of photons reaching the respective detectors during the sampling interval. Assuming that (i) photons are generated by a spatially nonhomogeneous Poisson process with intensity X_s per time unit, and (ii) the orientations θ of photon trajectories are uniformly distributed on $[0, 2\pi)$, we can show that Y_t, for $t \in D = \cup_{k=1}^{n} D_k$, is itself a Poisson process with a nonhomogeneous intensity function described in terms of the attenuated Radon transform (ART) of X. The ART of X is defined as

$$(R_{\mu,T}X)(\sigma,\theta) = \int_{\mathcal{L}} TX(x,y)\exp(-\int_{\mathcal{L}(x,y)} \mu(\xi,\eta)dl')dl$$

where $\mathcal{L}$ is the line with orientation θ, through point σ of the detector array, $\mathcal{L}(x, y)$ is the segment of $\mathcal{L}$ starting at point (x, y) in Ω, and dl and dl' are differential arc length in the two line integrals. The intensity function of Y is then given by

$$EY_t = \int_{\theta_k-\Delta\theta/2}^{\theta_k+\Delta\theta/2} \int_{\sigma_j-\Delta\sigma/2}^{\sigma_j+\Delta\sigma/2} (R_{\mu,T}X)(\sigma,\theta)d\sigma d\theta,$$

where $t = (\sigma_j, \theta_k)$. The important feature of this representation is that the intensity function of Y is the result of applying a positive linear integral operator $\mathcal{A}_T$ to X:

$$EY = \mathcal{A}_T X. \tag{5.3}$$

By further exploiting properties of the Poisson process, it can be shown that the observables Y_t are mutually independent and Poisson distributed; the likelihood function

is then easily obtained from (5.3). Recall our discretization of the domain Ω into *pixels* parameterized by discrete points s in a square lattice S. Now $\{X_s\}_{s\in S}$ represents a piecewise constant approximation of the isotope concentration on the continuous domain. When Ω is discretized, then equation (5.3) takes the form

$$EY = A_T X,$$

where A_T is a matrix, $A_T = \{A(t,s)\}_{t\in D, s\in S}$; commonly, the order of A_T is extremely large and it may not have full column rank. Now for a given X, the Poisson probability function of Y is

$$\Pi(Y|X) = \prod_{t\in D} \frac{[(A_T X)(t)]^{Y_t}}{Y_t!} \exp\{-(A_T X)(t)\} \tag{5.4}$$

It is interesting to examine reconstruction by the method of maximum likelihood, which calls for choosing X to maximize $\Pi(Y|X)$, given observed counts Y. Shepp and Vardi [55] have developed effective algorithms based on EM (Dempster, Laird and Rubin [16]) for implementing ML reconstructions in positron emission tomography (PET). A penetrating description, written from a statistician's perspective, is given in Vardi, Shepp and Kaufman [57]. (In PET, photon attenuation does not enter the model relating isotope concentration to the observables.) McClure and Accomando [1] have developed the foundations for applying ML to SPECT reconstructions and have implemented EM algorithms on a variety of computer systems. Independently, Miller, Snyder and Miller [48] have made similar extensions of ML and EM for SPECT.

The log-likelihood function is

$$\ln L(X) = \sum_{t\in D} \{-\ln(Y_t!) + Y_t \ln[(A_T X)(t)] - (A_T X)(t)\}. \tag{3.2}$$

The necessary conditions for maximizing $\ln L(X)$ obtained by setting derivatives to zero do not yield explicit solutions for a maximizing X. Nonetheless, $-\ln L(X)$ is globally convex, and the ML optimization problem conveniently adapts to the EM method. In general, $-\ln L(X)$ is not *strictly* convex; this is an identifiability issue related to the column rank of A_T. Conditions for *strict* convexity are discussed by Accomando [1].

POSTERIOR DISTRIBUTION. From (2.1),(5.1),(5.2), and (5.4) the *posterior distribution* on X is

$$\Pi(X|Y) = \frac{1}{Z(Y)} \exp\{-U(X) + \sum_{t\in D} [Y_t \ln[(A_T X)(t)] - (A_T X)(t)]\} \tag{5.5}$$

where $Z(Y)$ is a normalizing constant that depends on Y.

ESTIMATING THE IMAGE. Our experiments indicate that the minimum mean squared error (MMSE) estimator ($X^* = E(X|Y)$) is more appropriate for the reconstruction problem than the MAP used for texture classification (§3) and boundary detection (§4).

COMPUTING. We initialize $X = X(0)$. In practice, we choose a "good" initialization, such as the maximum likelihood reconstruction via EM, but easy theory says that convergence is independent of the initialization. Stochastic relaxation produces an *ergodic* Markov chain with equilibrium (5.5). The ergodicity of the chain guarantees that an ergodic average of $\{X(\tau)\}_{\tau=0}^{\infty}$ will converge to X^* a.s. In practice, we compute N iterates and average the final M, with choices such as $N = 25$ and $M = 5$. The selection of suitable M and N can be guided by monitoring stabilization of statistics of the successive iterates $X^{(\tau)}$.

PARAMETER ESTIMATION. The choice of β is critical. With $\beta = 0$ the estimator is undersmoothed, and in fact MAP estimation is just ML, since the prior is uniform. If β is too large, the estimator is too faithful to the prior and is oversmoothed. The parameter δ is also important, though we have found that (i) its value can usually be set based on information about the range of values $\{X_s\}$, and (ii) reconstructions are not sensitive to moderate changes in δ. The discussion here will focus on β.

Because of the setting in which reconstruction algorithms are actually used, it is desirable to design estimation methods that work with a sample $\tilde{Y}$ of size one from the observable process. The isotope density $\tilde{X}$ is assumed to be drawn from a Gibbs prior with unknown β, but known δ (5.2). We shall estimate β from $\tilde{Y}$ and use the estimate $\hat{\beta}$ in the MMSE or MAP reconstruction program. It is reasonable to do this with a single observation $\tilde{Y}$, since $\tilde{Y}$ contains a large amount of data about $\tilde{X}$, which, in turn, contains a large amount of data about the local energy function $U(X)$.

To be more explicit about the dependency on β of the prior and posterior distributions, we introduce the function

$$V(X) = \sum_{\{s,t\}} \phi(X_s - X_t) + \frac{1}{\sqrt{2}} \sum_{<s,t>} \phi(X_s - X_t).$$

V is just U/β. The prior is now written

$$\Pi(X) = \frac{1}{Z_\beta} \exp\{-\beta V(X)\}$$

and the posterior, given $\tilde{Y}$, is

$$\Pi(X|\tilde{Y}) = \frac{1}{Z_\beta(\tilde{Y})} \exp\Big\{-\beta V(X) + \sum_{t \in D} \big[\tilde{Y}_t \ln[(A_T X)(t)] - (A_T X)(t)\big]\Big\}$$

Now $V(X)$ is a *complete-data sufficient statistic* for β. If we were able to observe $\tilde{X}$ directly, then we could, in principle, solve the likelihood equation

$$E_\beta[V(X)] = V(\tilde{X}) \tag{5.6}$$

for the ML estimate of β. The left-hand side of (5.6) is strictly decreasing in β and thus (5.6) yields a unique root $\hat{\beta}$.

Our situation is more complicated than this since we do not observe $\tilde{X}$, but instead we see only the *incomplete data* $\tilde{Y}$. We have a classic setup for application of EM. The EM algorithm, when it converges, will yield a root of the incomplete-data likelihood equation

$$E_\beta[V(X)] = E_\beta(V(X)|\tilde{Y}); \tag{5.7}$$

see Dempster, Laird and Rubin [16]. We note that there is no proof of uniqueness of roots of (5.7). Conceptually, (5.7) is solved at the intersection of two monotone decreasing functions of β. Whether (5.7) does admit multiple solutions is an open and elusive theoretical question.

To solve (5.7), the EM algorithm consists of two alternating steps—estimation of the right-hand side of (5.7) for prescribed β (E-step) and computation of the root $\tilde{\beta}$ of (5.7), substituting the current estimate of $E_\beta(V(X)|\tilde{Y})$ on the right-hand side. Specifically, we fix an initial $\beta = \beta^0$ and an initial $X = X^0$ (and hence V^0). Then solve

E-step. Estimate the complete-data sufficient statistic:

$$V^{(\tau+1)} = E_{\beta^{(\tau)}}(V(X)|\tilde{Y}) \tag{5.8a}$$

M-step. Determine $\beta^{(\tau+1)}$ as the solution of

$$E_\beta[V(X)] = V^{(\tau+1)}. \tag{5.8b}$$

The first step is done using stochastic relaxation, using say ten steps of stochastic relaxation and averaging the last five values of $V(X^{(\nu)})$. The second step is a simple root-finding calculation once the curve $E_\beta[V(X)]$ is known. Conveniently, the stochastic relaxation procedure simultaneously yields updates $X^{(\tau)}$ of the MMSE reconstruction. Thus (5.8a) and (5.8b) together give a completely data-driven method of reconstruction.

The construction of $E_\beta[V(X)]$ as a function of β can be done off-line, once and for all. We have done this using stochastic relaxation to simulate 230 configurations X from the prior (4.1) for β-values ranging from 0 to 6. Five replications were done at each of forty-six values of β. The resulting curve, fit by a cubic-spline regression function, is depicted in Figure 15. The calculation of this curve required forty-one hours of CPU time, using a highly optimized program on the 100 Megaflop Star Technologies ST100 Array Processor.

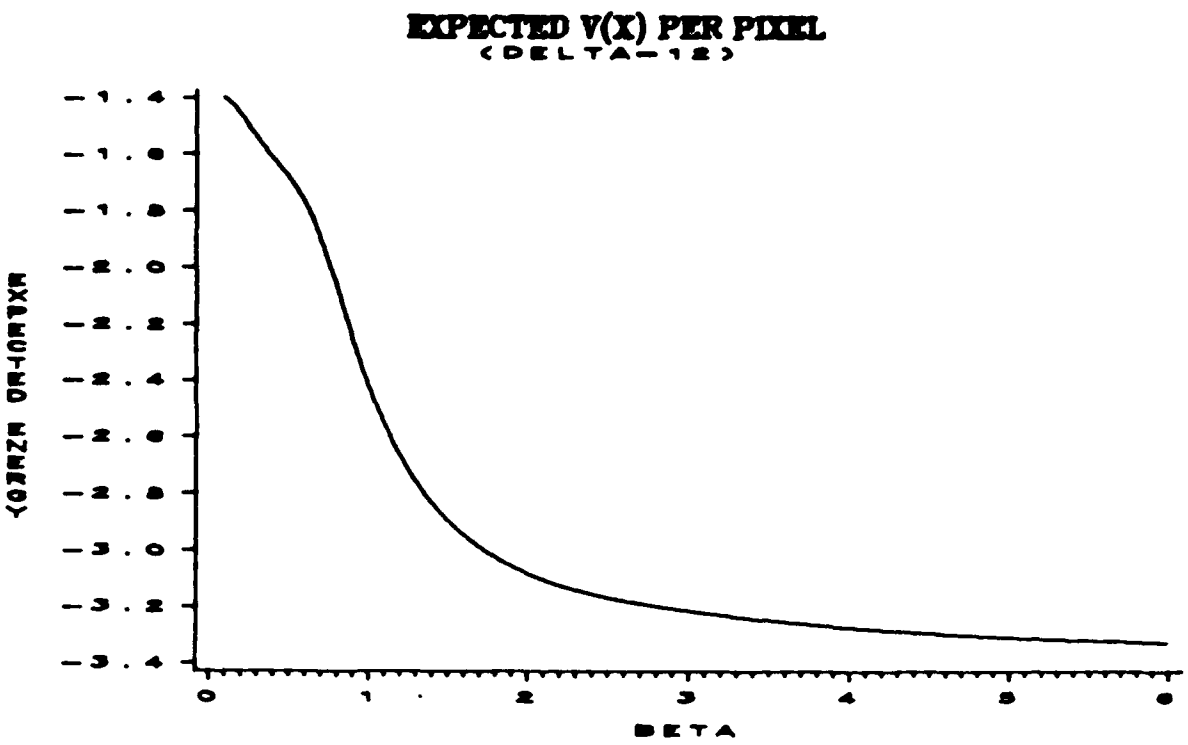


Figure 15.

J. Mertus [46] has developed an efficient vectorized FORTRAN program for the EM estimation/reconstruction procedure described above. Each E-step, with ten sweeps of stochastic relaxation, takes on the order of seven minutes of CPU time on an IBM3090 or about three minutes on a CYBER 205, working on a 64 x 64 pixel lattice S, for isotope densities X having their support on a disk of diameter 44 pixels (about 22cm) and with a range of 64 grey levels. (These values correspond to our real data sets.) The computational requirements are enormous, but not prohibitive.

EXPERIMENTAL RESULTS. We report on two experiments which have been run on real and simulated data to learn about the performance of the Bayesian reconstruction

methods. One simulation experiment was designed to test the versatility and robustness of the methods to known departures from the model. The other experiment illustrates the performance of the algorithms on real data from a lung section.

Experiment 1. A phantom isotope density (Figure 16A) was designed to have a combination of (i) large-scale structure, including subregions of Ω with considerable differences in intensity, and (ii) local irregularity of the same qualitative nature as that of sample functions from the Gibbs model with energy (5.1)–(5.2), yet not precisely fitting the Gibbs model. Two functions were averaged to form the phantom. First, an array with a sharp spike in intensity (near the center, below the middle) was constructed. Second, an array was sampled from (2.1) (using (5.1)–(5.2)) with parameter values $\beta = 1$, and $\delta = 12$. Intuitively, the local structure of the average will be governed by the array sampled from the Gibbs model. But observe that the rescaling of this array due to the arithmetic averaging means that it will not exactly fit a model from the same family. Roughly speaking, the averaging has the effect of smoothing the array so that it will be better described by a Gibbs model with larger β-value, assuming δ is fixed for now. We thus anticipate estimated values of β larger than the value $\beta = 1$ used to generate the Gibbsian part of the averaged phantom.

To simulate the emitted photons, the constant linear attenuation function $\mu \equiv 0.2$ was chosen, corresponding to approximately ten percent attenuation per centimeter for our scaling of the real system. A total of 663,144 photons were counted at 64 angles θ, with $L = 64$ bins on the linear detector array; in actuality, only 44 of the bins collect positive counts because the support of the phantom is contained in a smaller disk of diameter 44 pixels.

Reconstructions are depicted in Panels B–E of Figure 16. All were constructed on the range $[0, 63]$ with parameter $\delta = 12$. The MMSE reconstruction, with β estimated by EM (5.8) is shown in Panel 16B. When β was initialized at $\beta^0 = 0.0$, the successive EM iterates from (5.8b) were 0.52, 0.72, 0.85, 1.01, 1.18, 1.29, 1.34, 1.38, 1.40, 1.41, ..., 1.47 after thirteen steps of (5.8b). The MMSE in Panel 16B was run at $\hat{\beta} = 1.47$.

The *maximum likelihood* (via EM) reconstruction is shown in Panel 16C. Panel 16D shows an MMSE run with a value of $\beta = 0.52$, which is too small (undersmoothing). Panel 16E shows an MMSE run with a value of $\beta = 4.40$, which is too large (oversmoothing).

FIGURE 16 IS PLACED AT THE END OF THE TEXT.

Experiment 2. A total of 124,136 photons were counted from a cross section of a patient's

torso, including the lungs. For the reconstructions, we set the linear attenuation function again at $\mu \equiv 0.2$. The reconstructions were done on the range $[0,63]$ with fixed $\delta = 12$.

Panel 18A shows the maximum likelihood (EM) reconstruction. The "hot spot" (tumor) in the lung is apparent, but local structure is difficult to distinguish. Panel 18B shows the MMSE reconstruction with β estimated at $\hat{\beta} = 4.56$ after four steps of the EM estimation procedure (5.8); here we initialized $\beta^0 = 6.0$.

FIGURE 18 IS PLACED AT THE END OF THE TEXT.

DIRECTIONS FOR FUTURE RESEARCH. ***(1) Fast, Direct Parameter Estimators.*** The smoothness of the reconstructions is sensitive to the pivotal parameter β. Controlled experiments, using samples from the Gibbs prior with *known* β, indicate that EM (via equation (5.8)) is an effective procedure for estimating β. However, it is very computationally intensive. We are continuing to study an alternative *method of moments* estimator. The goal is to have a direct estimation method for β that can be applied to the observable $\tilde{Y}$ without requiring intermediate reconstruction of X. We construct a statistic $M(Y)$ based on the notion that the smoothness of Y will reflect the magnitude of β in the same way that the smoothness of X does.

As an example consider the following statistic, which we intend to study further. For the detector bin at angle θ_k and at sampling step σ_j, denote $t = (\sigma_j, \theta_k)$ and $t^+ = (\sigma_{j+1}, \theta_k), j = 1, \ldots, L-1,\ k = 1, \ldots, n$. Also, introduce the notation $a(t) = (A_T 1)(t)$, where 1 is the vector with components identically equal to one; $a(t)$ is simply the row-sum of A_T associated with the detector at location t. Then define the moment statistic

$$M(Y) = \sum_{k=1}^{n} \sum_{j=(L/2)-R}^{(L/2)+R} \left[\left(\frac{Y_t}{a(t)} - \frac{Y_{t^+}}{a(t^+)} \right)^2 - \frac{Y_t}{a^2(t)} - \frac{Y_{t^+}}{a^2(t^+)} \right]. \tag{5.9}$$

In the limits on the summation R is a *radius* of a window on the detector array over which the summands are accumulated. The expectation of $M(Y)$, for given X, is a measure of roughness of normalized ART projections of X:

$$E(M(Y)|X) = \sum\sum T^2 \left[\frac{A_T X(t)}{a(t)} - \frac{A_T X(t^+)}{a(t^+)} \right]^2.$$

We anticipate that the expectation $E_\beta[M(Y)]$ with respect to the prior will have the same general behavior as $E_\beta[V(X)]$ in (5.6). Accordingly, we define the moment estimate β^* of β as the root of the equation

$$E_\beta[M(Y)] = M(\tilde{Y}). \tag{5.10}$$

The effort to compute β^* is trivial, once the left-hand side of (5.10) is known as a function of β.

We have constructed the curve describing $E_\beta[M(Y)]$ using the same simulated X-data that generated $E_\beta[V(X)]$ in Figure 15. Figure 17 shows the resulting curve; it does, indeed, exhibit the same qualitative behavior as the curve in Figure 15. We plan to test this and other moment estimators on real and simulated data, in order to develop an accurate and computationally effecient estimator for β.

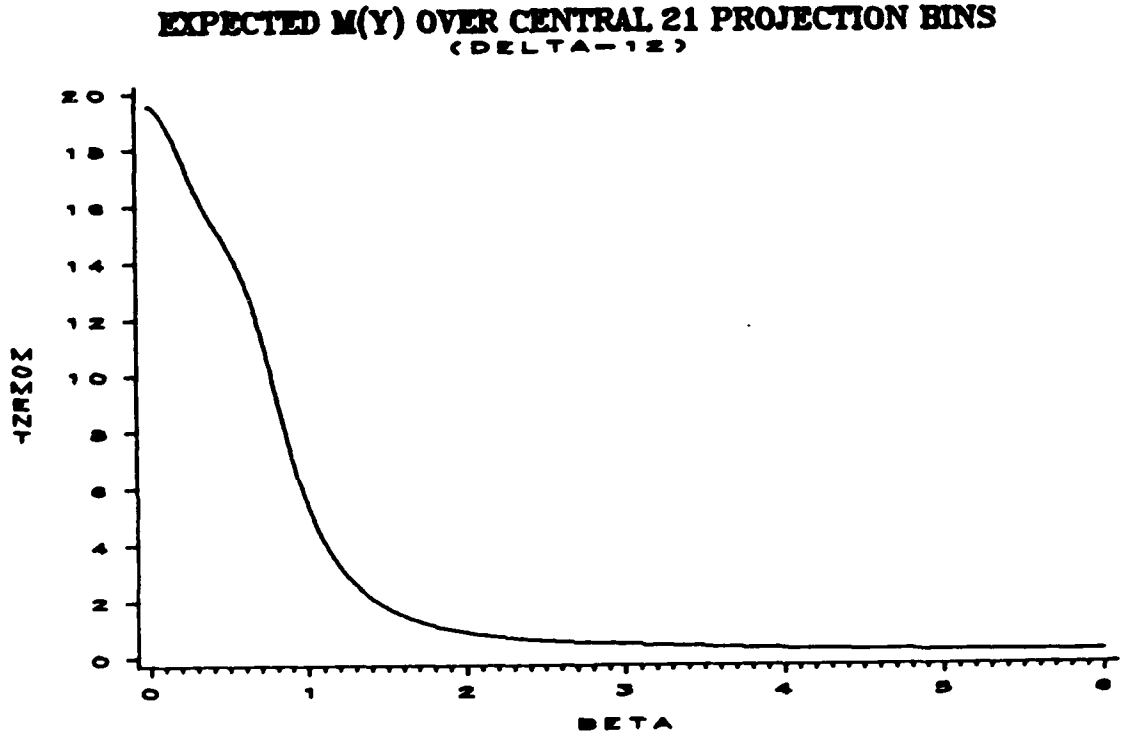


Figure 17.

Our initial experiments with the moment estimator suggest that the statistic $M(\cdot)$ is not robust as one would like with respect to certain kinds of boundary singularities in X. We intend to study alternatives to M in which the summands are less sensitive to large local variations than are the quadratic terms in (5.9). Analytical difficulties are compounded, however, for the robust alternatives.

(2) Improvements of the Degradation Model. The model presented here includes the predominant physical effects. But we have assumed that other potentially significant effects,

such as photon scattering and background radiation, to be negligible. Further, we have not included effects from the sensor, such as imperfect collimation, blurring, and noise. It has been established that these effects can contribute significantly to the degradation (see, for example, [19]). We note, however, that the reconstruction methods described, since they are based on the generally applicable principles of maximum likelihood and Bayes optimality, are adaptable to models incorporating additional physical and sensor effects. We are currently studying extensions for scattering and collimation errors. These extensions will significantly increase the effective size of the ART, and thereby present difficult computational challenges.

(3) Alternative Methods of Noninvasive Testing. We intend to study extensions of our Bayesian reconstruction method to other modalities of noninvasive imaging and nondestructive testing. Currently we are initiating studies with ultrasonic imagery.

6. GLOBAL IMAGE ANALYSIS

Our model-based approach to image processing has its roots in a general pattern theory layed out by Grenander [32]. In our discussion here of global image analysis we shall adopt the terminology and notation developed for that theory, which is especially well suited for describing global geometric attributes. There is an equivalent, but less natural, formulation in the statistical mechanics/MRF language used in the discussion above.

AIM. Our approach is *model based.* Starting from a mathematical model of the patterns (image ensemble) that we are dealing with we *derive* algorithms for pattern inference from the models by appealing to first principles.

We are primarily concerned with patterns exhibiting a high degree of variability, such as those appearing in biology/medicine. It is then natural to account for the variability by probabilistic models.

A situation of obvious practical importance, as well as scientifically challenging, is when we know in advance that the image contains an object of known type but of unknown shape. We may know for example that the image contains a kidney, a heart, a leaf, and so on, and we seek a precise characterization of the object: *object identification.*

The purpose may be to detect anomalies or to measure certain characteristics on the globally estimated object. What matters then is the variability in the object shape.

We desire an accurate characterization even though the picture may be very noisy. To get that we appeal to the prior knowledge we have of the variation in shape for the object.

It is clear that global models are needed for such a task, it is a matter of *high level image analysis.*

GENERAL APPROACH. We have long employed methods from metric pattern theory to formulate and analyze the random variation of patterns. An early realization of this idea was given in Grenander [30], and was applied to pictorial patterns in Grenander [31], pp. 40-51.

The resulting probability measures are known as regularity controlled probabilities, or Gibbs measures on configuration spaces, or incompletely observed Markov processes on graphs. The underlying principle is that they attempt to catch variability by relaxing regularity in the patterns.

Recently this was generalized and studied in greater depth in Grenander-Keenan [34] in the following way. Starting from primitive geometric objects, the generators g, in some space G, we combine them into configurations

$$c = \sigma(g_0, g_1, \ldots, g_{n-1}) \in C(\mathcal{R}) \tag{6.1}$$

in a configuration space. Here σ denotes a connector graph with n sites, and regularity constraints are imposed on any generators connected directly by a segment in the graph; the generator g_i is situated at site i of σ.

We start with one or several templates $c^{(\nu)}$, not necessarily with the same connector graph. For a given group S, the similarity group, with elements generically denoted by s, we define a measure on S^n by

$$p(s_0, s_1, \ldots, s_{n-1}) = \frac{1}{Z} \prod_{i=0}^{n-1} A_i(s_i, s_{i+1}) \prod_{i=0}^{n-1} Q_i(s_i) \tag{6.2}$$

where p is understood as a Radon-Nikodym derivative with respect to a fixed measure on S^n , usually a Lebesgue or counting measure. The acceptor functions A_i express the coupling between contiguous group elements s_i—group elements that are compatible with the connector graph σ—and the weight functions Q_i control the frequencies of their occurrence.

Applying the stochastic operators to the templates

$$c^{(\nu)} \to (s_0, s_1, \ldots, s_{n-1})c^{(\nu)}, \tag{5.8}$$

but restricted to the regular configuration space $C(\mathcal{R})$, will define a probability measure on $C(\mathcal{R})$. A rigorous study of how this is done can be found in Grenander-Keenan [34] and in Grenander [32] (Volume III, Chapter 3).

To make the above concrete let us consider some examples. For pictorial patterns in the plane we can choose, for example, generators as points in R^2 and S as the translation group. Or we can let the generators be sides in polygons (or splines with other arc elements) and S the group of uniform scale change. Or, with the same generators, we can let S be the affine group in the plane, and so on.

In these examples the obvious choice for the connector graph is $\sigma \in CYCLIC$, but alternatives with 'transversal' bonds are also of interest for generating patterns with (relaxed) symmetry, for example of bilateral form.

We believe that the creation and mathematical analysis of such random shape models are fundamental for achieving the tasks of global image analysis. Once the models are well understood and empirically supported we can apply methods of statistical inference to them.

SPECIAL RESULTS. We have carried out pattern inference in 2-D as outlined above in a number of instances.

Simulated Pictures. To test the methodology, image reconstruction was implemented based on random shape models. The result compare favorably with local methods. Some examples are given in Grenander [33] and Grenander-Keenan [34].

Leaf Shapes. Simulated data are good for testing methods and software, but are not enough. It is necessary to try the methods on real data. We have therefore also applied the same methodology to a type of biological patterns, namely the shape of leaves from trees. The data collection phase has been completed, as well as the development of software for handling images obtained by hardware we have acquired recently. Algorithms, both for image restoration and object identification, have been derived and some preliminary experiments have been carried out. This work with A. Knoerr is still in progress.

Limit Theorems. We have used our method of stochastic relaxation extensively in our computer experiments. The power of it resides in its general applicability, but we pay a price for its universality: it often requires massive amounts of CPU time. This is especially true in 3-D; see below.

We have tried to compensate for this by analytical means. It has been known for several years that, in special cases, and when stochastic relaxation is slowest (strong couplings, large configurations), the Markov processes on graphs can be approximated by Gaussian processes; see Chow-Grenander [13].

More recently it has been shown that such limit theorems can be proved in more

general situations, Chow [12] and Grenander-Sethuraman [35]. Since Gaussian processes are easily simulated this makes it possible to reduce the CPU-time requirement drastically.

Fitting the Models. In order that our analyses be practical the model must be anchored empirically. In particular their parameters must be estimated from data. In the random shape models we do not have statistical homogeneity: generators at different sites may have quite different statistical properties. We must therefore study estimation based on a sample of several pictures. This has been done in Osborn [51].

3-D Experiments. We have tried our models in 3-D situations, so far only for pattern synthesis. In this case we used stochastic relaxation in a straightforward manner, since we do not yet have access to the sort of limit theorems for 3-D mentioned above. The computing is so heavy that it was necessary to use a supercomputer, a CRAY XMP 48.

DIRECTIONS FOR FUTURE RESEARCH. Based on the methodology we have developed and the the experience we have gained in applying them, we shall proceed to more challenging tasks in global image analysis.

With priors of the form (6.2) and denoting the resulting random image by I we relate I to the observed digital $I^{\mathcal{D}}$ by a stochastic deformation mechanism $\mathcal{D}$

$$\mathcal{D} : \mathcal{I} \to \mathcal{I}^{\mathcal{D}}$$

mapping the pure image algebra $\mathcal{I} = R\mathcal{I}$ (where R is the identification rule, see Grenander [32], Volume I, Secton 3.1) into the deformed one. Before discussing the form of $\mathcal{D}$ it is necessary to mention some difficulties that arise as we go from 2-D to 3-D.

As generators we can choose vertices or edges or faces of polyhedra (or other spline surfaces): all three possibilities will be examined. We shall choose the connector graph s for the template (or templates) as the vertices of a regular icosahedron further refined by successive triangulations, using equilaterals. This results in a regular distribution of points on the template surface assumed to be topologically equivalent to the sphere, and avoids the singular behavior of σ at the poles if we had used a standard longitude-latitude grid. The successive triangulations are useful for implementing the inference algorithms which will be hierarchic: starting with crude templates they will be successively refined.

Global 3-D Inference. Extending what we did in R^2 to R^3 we have to simulate the posterior

$$p(c \mid I^{\mathcal{D}}) \propto p(c)p(I = Rc \to I^{\mathcal{D}}) \tag{6.5}$$

where the last factor is due to $\mathcal{D}$.

We shall investigate three models for the deformation $\mathcal{D}$:

a) visible light

b) range data

c) ultrasound.

We already have the hardware and some of the software for a). At present we do not have any sensors for b) and c) and will have to operate with simulated data during the stage when we fix the models and derive inference algorithms.

This is not sufficient, however, we should also test the methods in a real world setting. We shall therefore try to get access either to sensors for b) and c), or, at least to data obtained under realistic conditions. We do have good sources for echocardiographic data (c), in particular.

To simulate (6.5) we shall try to avoid straightforward stochastic relaxation with single site updating since this would be computationally awkward. Instead we appeal to pattern theoretic limit theorems of the type discussed above, so that we can simulate conditional probabilities for subsets of the configuration c with more than a single site. As a matter of fact we may be able to simulate the whole configuration using (6.2) combined with the factor $p(c \rightarrow I^{\mathcal{D}})$. This appears to be computationally feasible.

Note that in this way we make *global inferences* about the image that incorporate prior knowledge about the whole surface. A number of analytical problems have to be resolved, however, before this ambitious goal can be achieved.

The Correspondence Problem. Among these we mention in particular the correspondence problem. Consider the boundary described by the configuration c and a boundary c^* obtained from $I^{\mathcal{D}}$. Say that c^* consists of a set of points z_ν, for $\nu = 1, 2, \ldots, N$, that have arisen from some points ζ_ν on c. If we knew the ζ_ν we could calculate, at least in principle, the likelihood that $\mathcal{D}$ carries the $\{\zeta_\nu\}$ into $\{z_\nu\}$.

The trouble is that *this correspondence is not known to us.* Instead we will have to treat the ζ_ν as nuisance parameters that have to be estimated during the inference. In 2-D this can be done with reasonable computing effort by a dynamic programming algorithm. In 3-D we do not yet have any fast algorithm for dealing with the correspondence problem.

This is for the case when the inference is based on an 'empirical boundary' c^*. If, instead, we use only contrast based inference, as is natural for low signal-to-noise ratios, the correspondence problem does not appear.

Generator Estimation. We do not avoid it entirely, though, since we meet it during an earlier phase when the template surface (or curve in 2-D) is estimated from data. Again we do not know the correspondence. We plan to deal with this difficulty by extending a method for *template, or generator, estimation* that we have developed, but not yet tested, in R^2.

The Icosahedral Process. Another analytical topic that has to be studied is what are the properties of the priors to be used in 3-D for the random template surface. To fix ideas say that we choose edges as generators. The measure in (6.2) can be viewed as a Markov process on the triangulated icosahedral graph taking values in a low dimensional Lie group S. Applying the s_i-elements to the edges g_i^0 of the template we must have conditions satisfied in order that $g_i = s_i g_i^0$ are consistent

$$L_{cycle} = \sum_{cycle} s_i g_i^0 = 0 \tag{6.6}$$

for all closed cycles of the graph σ. We can find μ independent cycles, where μ is the cyclomatic number for the graph, by a constructive procedure starting from a minimal spanning tree (see for example, Giblin [26]). But, since the number of constraints will be large in 3-D in contrast to 2-D, we need a systematic and computationally feasible way of constraining the probability measure of the g_i's.

We should also extend the limit theorems for Markov processes on graphs described above to this sort of graph. The limiting Gaussian processes are believed to have covariance functions that are Green's functions of differential operators simply related to the neighborhood structure of the graph, but this statement should be made precise and proved.

Ergodicity. In situations where the limit theorems do not apply (for example with weak couplings) we are forced to employ stochastic relaxation. To justify stochastic relaxation we have to prove ergodicity (in time) of a Markov chain with a continuous state space. This reduces, in each special case, to an "elementary" problem in plane geometry.

As an example consider the configuration space C of all closed and non-self-intersecting polygons in R^2 (similarly in R^3) whose sides have given directions but arbitrary lengths. Denote the lengths by $l_1, l_2, \ldots, l_n$ so that we can represent a configuration as $c = (l_1, l_2, \ldots, l_n)$. The ergodicity mentioned can be established by proving that any two points

$$c_1, c_2 \in \text{interior}(C) \subset R_+^{n-2} \tag{6.7}$$

can be joined by a continuous curve passing through the interior of C.

This innocuous proposition has not yet been proved in spite of several attempts and discussions with a number of geometers. Many versions of it appear, for different updating schemes, but only some have been proved so far.

Software. We need software, for the object identification algorithms, but also for the data analysis of images that will precede and accompany the analysis of the models to be used.

During the data analysis stage computing efficiency is not as crucial as for the inference algorithms. We have started to develop data analysis programs in APL and plan to do this for the following tasks.

a) local Karhunen-Loève analysis

b) smoothing programs, linear, median, and trimmed mean filtering

c) primitive operations on surfaces: rotate around an axis, stretch, turn around an axis, push and pull the surface at a point, all for icosahedral/triangulation vertices

d) display template surfaces, I , and the object identified by our inference algorithms.

e) simulation of range data and ultrasound data

f) standard image operations like edge detection.

This code should be easy to write and will be useful for data analysis.

FIGURES.

Figure Legends.

Figure 2. Wood on plastic background.

Figure 3. Carpet on plastic background.

Figure 4. Wood, carpet, and cloth on plastic background.

Figure 10. Synthetic aperture radar of water and ice.

Figure 11. Brodatz collage with four textures.

Figure 12. House scene.

Figure 16. Phantom isotope density and reconstructions.

Figure 18. Lung cross-section and reconstructions.

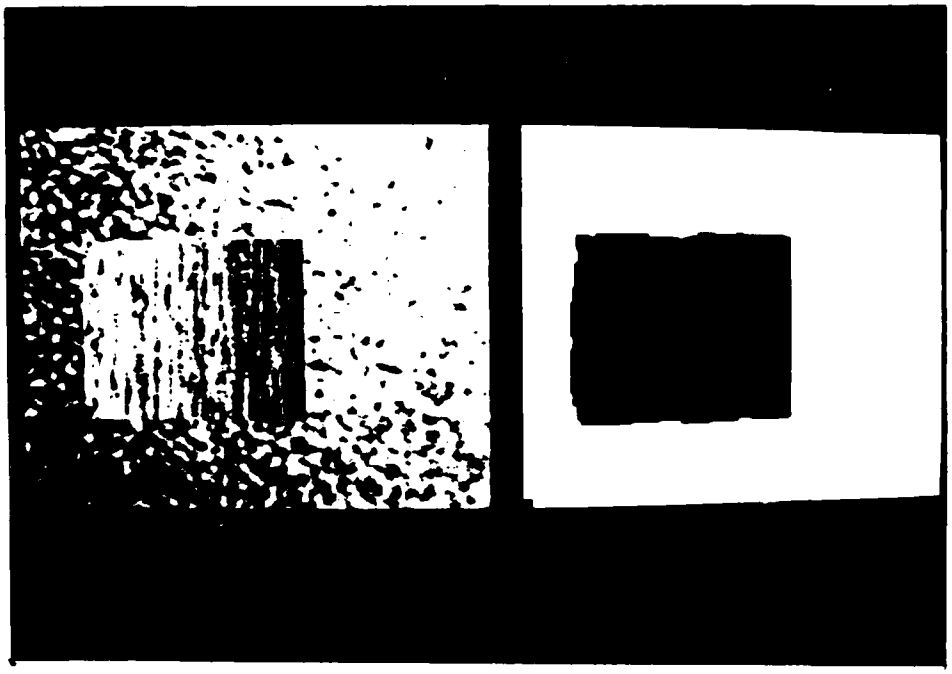


Figure 2. Wood on plastic background.

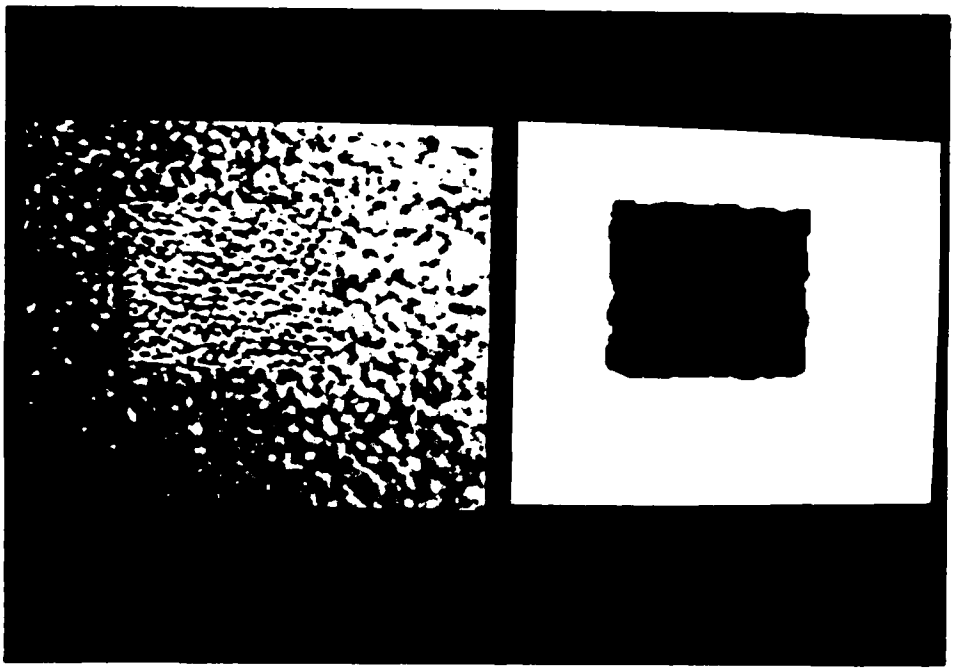


Figure 3. Carpet on plastic background.

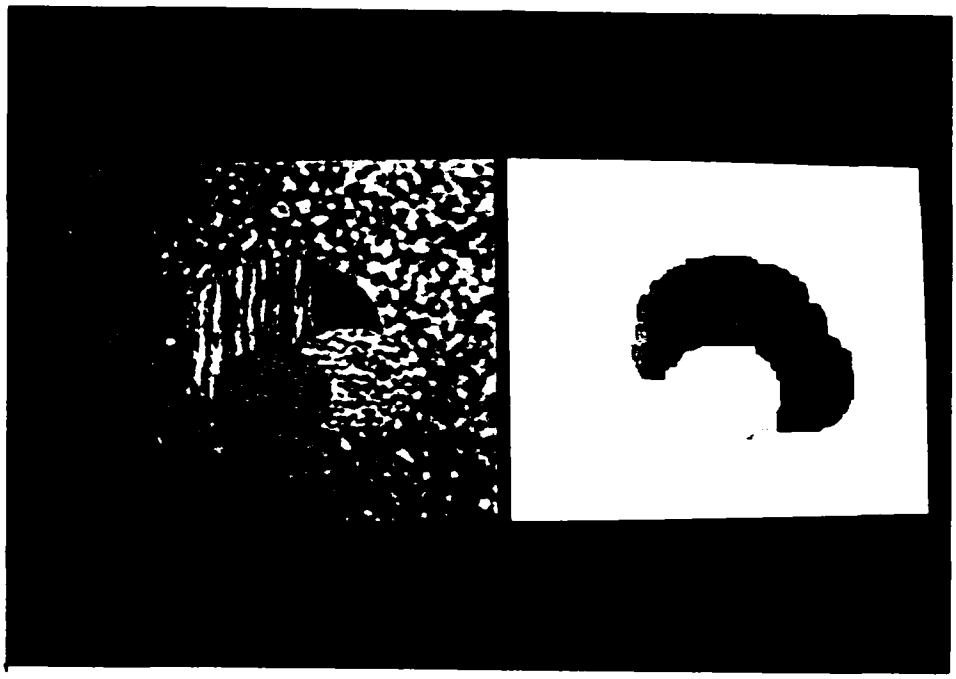
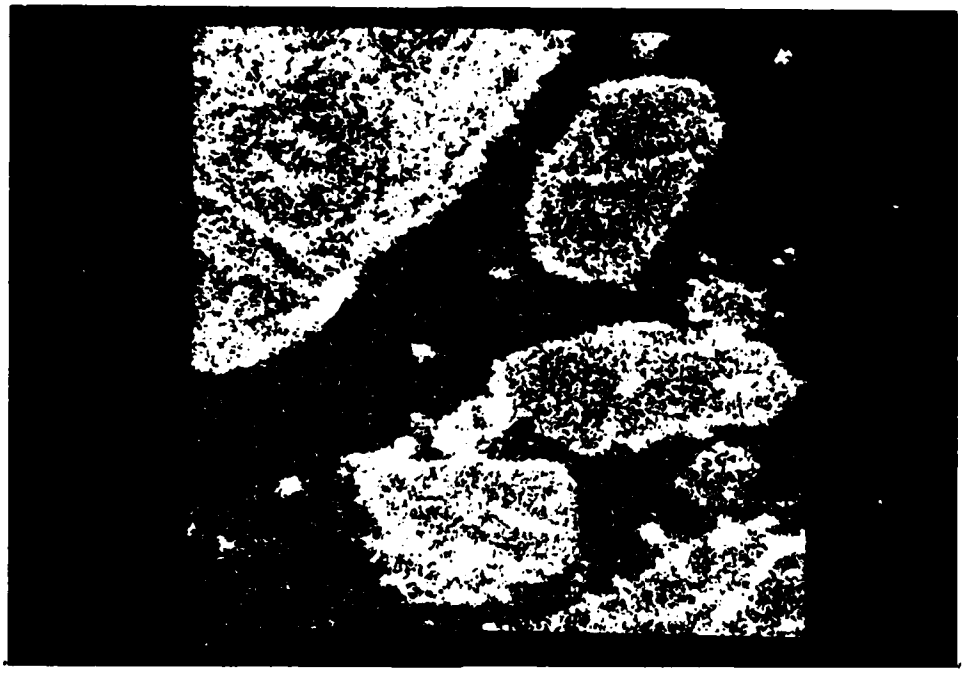


Figure 4. Wood, carpet, and cloth on plastic background.



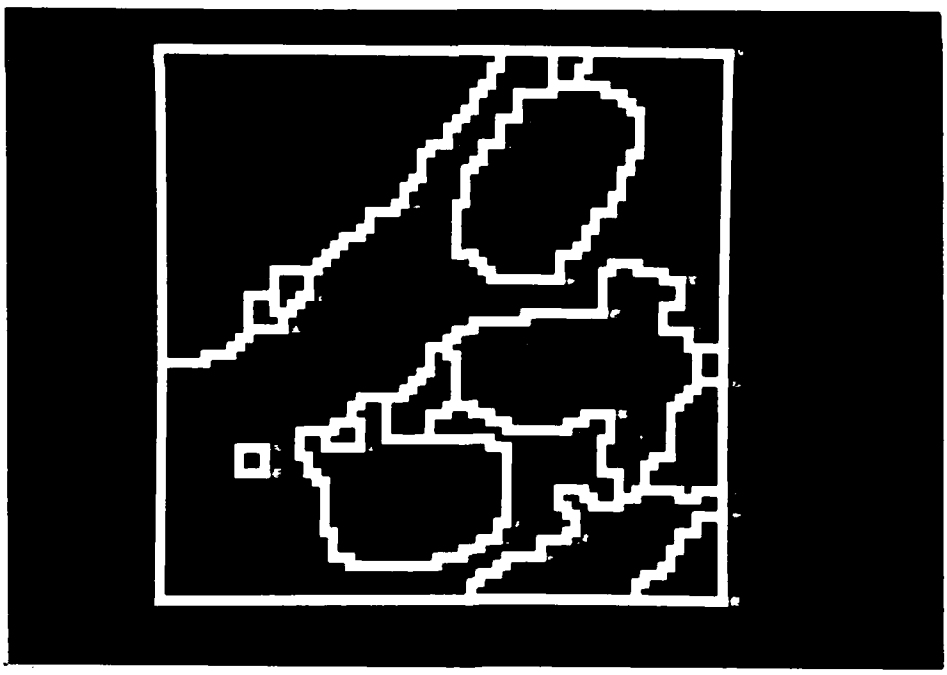


Figure 10. Synthetic aperture radar of water and ice.

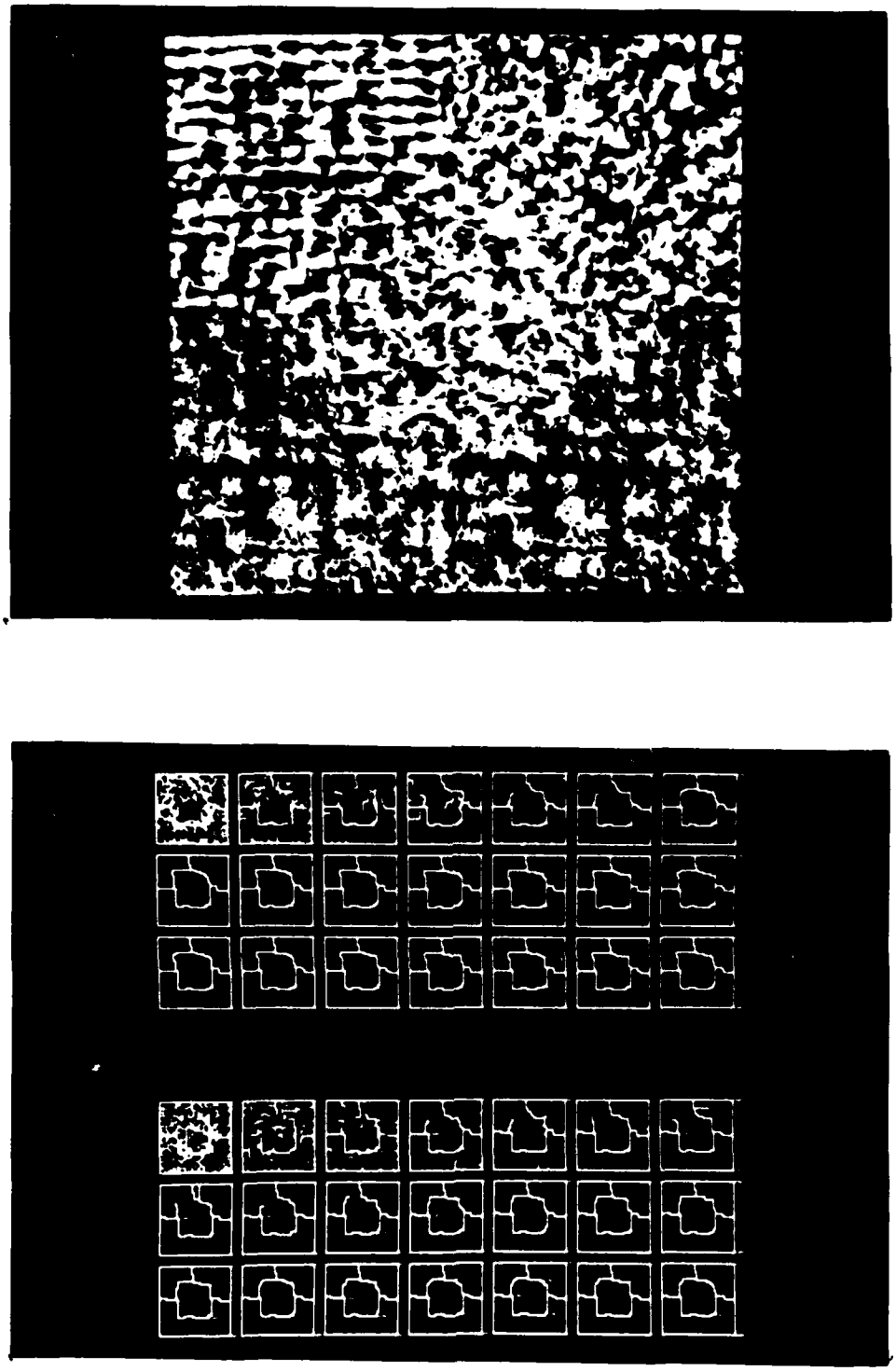
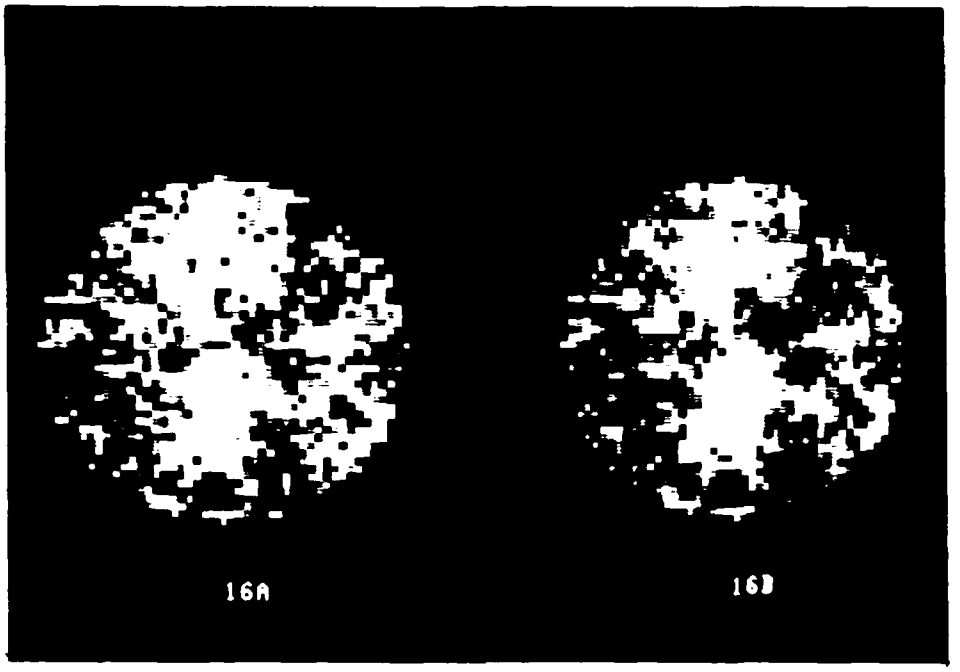


Figure 11. Brodatz collage with four textures.

Figure 12. House scene.



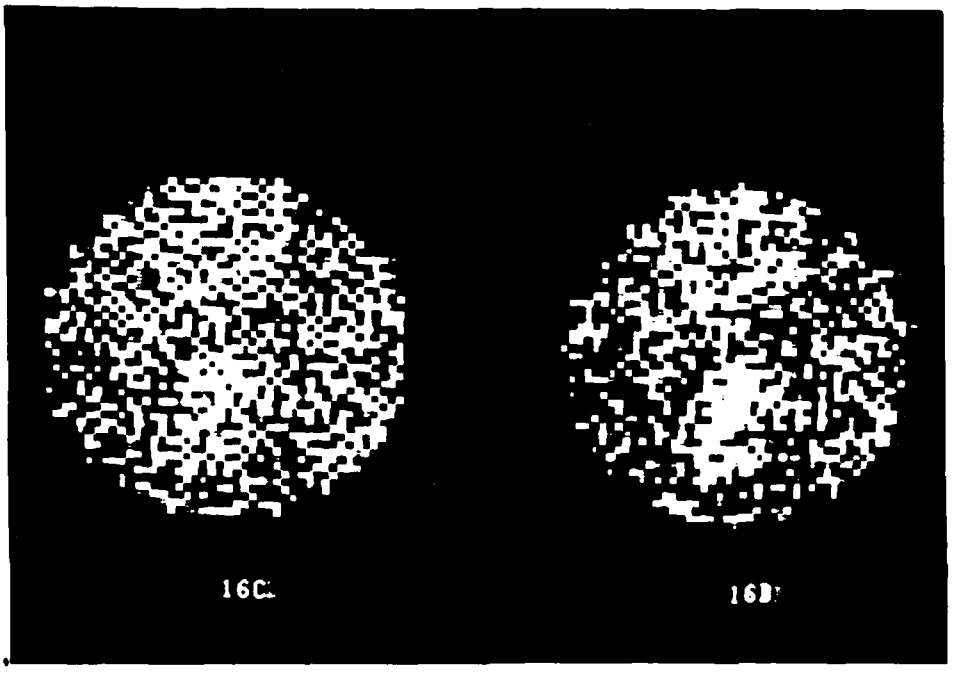


Figure 16A-D. Phantom isotope density and reconstructions.

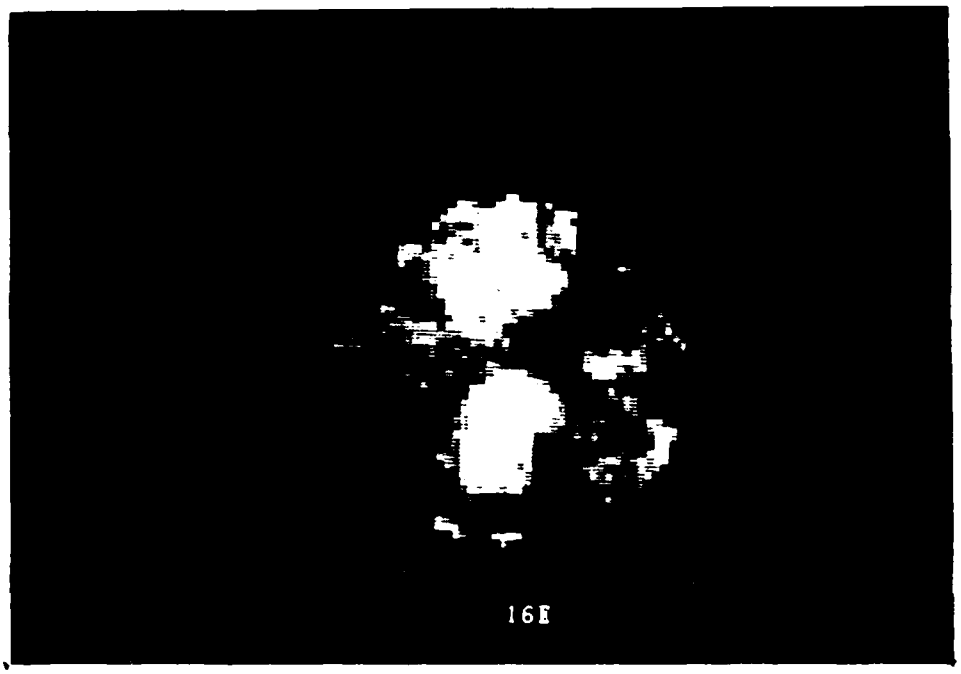


Figure 16E. Phantom isotope density and reconstructions (cont).

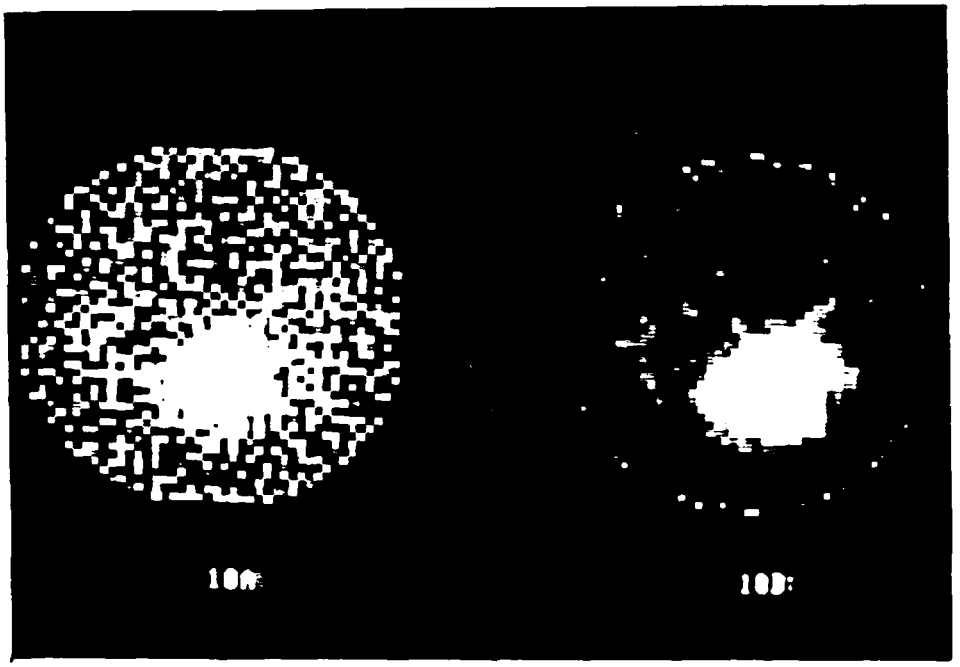


Figure 18. Lung cross-section and reconstructions.

BIBLIOGRAPHY

1. N. Accomando (1984), "Maximum likelihood reconstruction of a two-dimensional Poisson intensity function from attenuated projections," Ph.D. Thesis, Division of Applied Mathematics, Brown University.

2. J. Besag (1974), "Spatial interaction and the statistical analysis of lattice systems," (with discussion), *J. Royal Statist. Soc.*, Series B, Vol. 36, 192-236.

3. J. Besag (1975), "Statistical analysis of non-lattice data," *The Statistician*, 24, 179-195.

4. J. Besag (1986), "On the statistical analysis of dirty pictures," (with discussion), *J. Royal Statist. Soc.*, Series B, Vol. 48, 259-302.

5. A. Blake and A. Zisserman (1986), "Weak continuity constraints in computer vision," Report CSR-197-86, Dept. of Comp. Sci., Edinburgh Univ.

6. J.B. Burns, A.R. Hanson, and E.M. Riseman (1986), "Extracting straight lines," *IEEE Trans. Pattern Anal. Machine Intell.*, 4, 425-455.

7. L. Breiman (1973), *Statistics: With a View Toward Applications*, Houghton Mifflin Company, Boston.

8. P. Brodatz (1966), *Textures: A Photographic Album For Artists And Designers*, Dover, New York.

9. J. Canny (1986), "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, 8, 679-698.

10. V. Cerný (1982), "A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm," Inst. Phys. & Biophys., Comenius Univ., Bratislava (preprint).

11. B. Chalmond (1986), "Image restoration using an estimated Markov model," Mathematics, University of Paris-Sud, Orsay (preprint).

12. Y.S. Chow (1987), "A limit theorem for pattern synthesis in image processing," Div. Appl. Math. Report, Brown University, Providence, R.I.

13. Y.S. Chow and U. Grenander (1982), "A singular perturbation problem," *J. Integral Eqns.*

14. F.S. Cohen and D.B. Cooper (1986), "Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, (to appear)

15. G.R. Cross and A.K. Jain (1983), "Markov random field texture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-5, 25-40.

16. A. Dempster, N. Laird, and D. Rubin (1977), "Maximum likelihood from incomplete data via the EM algorithm," *J.R. Statist. Soc.*, Series B, Vol. 39, 1-38.

17. H. Derin and H. Elliott (1987), "Modelling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-9, 39-55.

18. H. Elliott and H. Derin (1987), "Modelling and segmentation of noisy and textured images using Gibbs random fields," Dep. Elec. Comp. Eng., Univ. of Massachusetts, Amherst, Tech. Rep. ECE-UMAS-SE84-15.

19. Carey E. Floyd, Jr., et al. (1984), "Energy and spatial distribution of multiple order Compton scatter in SPECT," *Phys. Med. Biol.*, 29:1217-1230.

20. D. Geman, S. Geman, and C. Graffigne (1986), "Locating texture and object boundaries," *Pattern Recognition Theory and Application*, Ed. P. Devijver, NATO ASI series, Springer-Verlag, Heidelberg.

21. D. Geman, S. Geman, C. Graffigne, and P. Dong (1987), "Locating surface boundaries by constrained optimization," (Preprint).

22. S. Geman and D. Geman (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, **6**, 721-741.

23. S. Geman and C. Graffigne (1987), "Markov random field image models and their applications to computer vision," In: *Proceedings of the International Congress of Mathematicians 1986*, Ed. A.M. Gleason, American Mathematical Society, Providence.

24. S. Geman and D.E. McClure (1985), "Bayesian image analysis: An application to single photon emission tomography," *Proc. American Statistical Association, Statistical Computing Section*, 12-18.

25. S. Geman and D.E. McClure (1987), "Statistical methods for tomographic image reconstruction," Proceedings of the 46th Session of the International Statistical Institute,

Bulletin of the ISI, Vol. 52.

26. P.J. Giblin (1977), *Graphs, Surfaces and Homology*, Chapman and Hall.

27. B. Gidas (1985), "Non-stationary Markov chains and convergence of the annealing algorithm," *J. Stat. Phys.*, **39**, 73-131.

28. B. Gidas (1986), "A renormalization group approach to image processing problems," Div. of Appl. Math., Brown University, Providence (Preprint).

29. D.M. Greig, B.T. Porteous, and A.H. Seheult (1986), Discussion of "On the statistical analysis of dirty pictures" by Julian Besag, *J. of the Royal Stat. Soc.*, Series B, Vol. 48, 259-302.

30. U. Grenander (1967), "Syntax-controlled probabilities," Reports in Pattern Analysis No. 2, Div. Appl. Math., Brown University.

31. U. Grenander (1969), "Foundations of pattern analysis," *Quart. Appl. Math.*, Vol. XXVII, 1-55.

32. U. Grenander (1976,1978,1981), *Lectures in Pattern Theory*, 3 volumes, Springer-Verlag.

33. U. Grenander (1983), "Tutorial in Pattern Theory," Div. Appl. Math., Brown University.

34. U. Grenander and D.M. Keenan (1986), "On the shape of plane images," Reports in Pattern Analysis No. 145, Div. Appl. Math., Brown University.

35. U. Grenander and J. Sethuraman (1987), "Limit theorems in metric pattern theory," unpublished manuscript.

36. W.E.L. Grimson and T. Pavlidis (1985), "Discontinuity detection for visual surface reconstruction," *CVGIP*, **30**, 316-330.

37. B. Hajek (1985), "Cooling schedules for optimal annealing," *Mathematics of Operation Research*, (to appear).

38. R.M. Haralick (1984), "Digital step edges from zero crossings of second directional derivatives," *IEEE Trans. Pattern Anal. Machine Intell.*, **6**.

39. G.E. Hinton and T.J. Sejnowski (1983), "Optimal perceptual inference," *Proc. IEEE Conf. Comput. Vision Pattern Recognition.*

40. S. Kirkpatrick, C.D. Gellatt, and M.P. Vecchi (1983), "Optimization by simulated annealing," *Science*, **220**, 671-680.

41. E. Levitan and G.T. Herman (1987), "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. on Medical Imaging*, to appear.

42. Z. Liang and H. Hart (1987), "Bayesian image processing of data from constrained source distributions-I: Nonvalued, uncorrelated and correlated constraints," *Bull. Math. Biol.*, Vol. 49, 51-74.

43. A. Lippman (1986), "A maximum entropy methods for expert system construction," Ph.D. Thesis, Div. Appl. Math., Brown University.

44. D. Marr and E. Hildreth (1980), "Theory of edge detectors," *Proc. Royal Soc., B*, **207**, 187-207.

45. J. Marroquin, S. Mitter, and T. Poggio (1987), "Probabilistic solution of ill-posed problems in computational vision," *J. Am. Stat. Assoc.*, Vol. 82, 76-89.

46. J. Mertus (1987), "Self-calibrating Bayesian methods for image reconstruction in emission tomography," Ph.D. Thesis, Div. Appl. Math., Brown University, in preparation.

47. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller (1953), "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, **21**, 1087-1091.

48. M.I. Miller, D.L. Snyder, and T.R. Miller (1985), "Maximum-likelihood reconstruction for single-photon emission computed tomography," *IEEE Trans. on Nuclear Science*, NS-23, 769-778.

49. D. Mumford and J. Shah (1986), "Boundary detection by minimizing functionals, I," Harvard University, preprint.

50. V.S. Nalwa and T.O. Binford (1986), "On detecting edges," *IEEE Trans. Pattern Anal. Machine Intell.*, **8**, 699-714.

51. B. Osborn (1986), "Parameter estimation in pattern theory," Ph.D. Thesis, Div. Appl. Math., Brown University.

52. T. Poggio, V. Torre, and C. Koch (1985), "Computational vision and regularization theory," *Nature*, Vol. 317, 314-319.

53. A. Possolo (1986), "Estimation of binary Markov random fields," Dept. of Statistics, Univ. of Washington, Seattle, preprint.

54. D. Ruelle (1978), *Thermodynamic Formalism*, Addison-Wesley.

55. L.A. Shepp and Y.Vardi (1982), "Maximum likelihood reconstruction in positron emission tomography," *IEEE Trans. on Medical Imaging*, 1, 113-122.

56. D. Terzopoulos (1986), "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, 8, 413-424.

57. Y. Vardi, L.A. Shepp, and L.Kaufman (1985), "A statistical model for positron emission tomography," *JASA*, Vol. 80, 8-20 and 34-37.

END

12-87

DTIC